

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2002, 09:22:17 ; Search time 10.85 Seconds

(Without alignments)  
780.606 Million cell updates/sec

Title: US-09-445-480a-2

Perfect score: 1287  
Sequence: 1 MALSQNAKESKGFVYMIW.....VDEGVKVPSPVRFKRPNC 231

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1089	84.6	199	1	EOST_ACTEO
2	295.5	23.0	2766	1	THYG_MOUSE
3	295	22.9	2769	1	THYG_BOVIN
4	294.5	22.9	2768	1	THYG_RAT
5	272	21.1	2768	1	THYG_HUMAN
6	245.5	19.1	1375	1	NID2_MOUSE
7	244	19.0	844	1	SAX_RANCA
8	236	18.3	1403	1	NID2_MOUSE
9	156	12.1	279	1	HG2A_MOUSE
10	154	12.0	1245	1	NID2_MOUSE
11	153	11.9	280	1	HG2A_RAT
12	153	11.9	1247	1	NID2_HUMAN
13	150	11.7	296	1	HG2A_HUMAN
14	132.5	10.3	258	1	IBP4_BOVIN
15	129.5	10.1	254	1	IBP4_MOUSE
16	128.5	10.0	254	1	IBP4_RAT
17	124.5	9.7	237	1	IBP4_SHEEP
18	122.5	9.5	258	1	IBP4_HUMAN
19	122.5	9.5	271	1	IBP5_PIG
20	122.5	9.5	272	1	IBP5_HUMAN
21	120.5	9.4	271	1	IBP5_MOUSE
22	118.5	9.2	271	1	IBP5_RAT
23	114	8.9	263	1	IBP1_BOVIN
24	114	8.9	1786	1	IBP1_MOUSE
25	113.5	8.8	240	1	IBP6_HUMAN
26	112	8.7	1786	1	IBP1_HUMAN
27	110	8.5	266	1	IBP3_PIG
28	110	8.5	291	1	IBP3_BOVIN
29	109.5	8.5	259	1	IBP1_HUMAN
30	109.5	8.5	4544	1	IBP1_HUMAN
31	109	8.5	291	1	IBP3_MOUSE
32	108.5	8.4	292	1	IBP3_RAT
33	107	8.3	226	1	IBP6_MOUSE

34	107	8.3	238	1	IBP6_MOUSE	P47880 mus musculus
35	291	8.3	291	1	IBP3_HUMAN	P17936 homo sapien
36	107	8.3	305	1	IBP2_MOUSE	P47877 mus musculus
37	107	8.3	1429	1	L112_CAEEL	P14585 caenorhabdi
38	106	8.2	272	1	IBP1_RAT	P21743 rattus norv
39	105.5	8.2	863	1	LDVR_CHICK	P98165 gallus gall
40	104	8.1	810	1	NEB1_HUMAN	O92832 homo sapien
41	103.5	8.0	304	1	IBP2_HUMAN	P12843 rattus norv
42	103	8.0	272	1	IBP1_MOUSE	P47876 mus musculus
43	102	7.9	311	1	IBP2_CHICK	P49705 gallus gall
44	100.5	7.8	314	1	G732_HUMAN	P16422 homo sapien
45	100	7.8	1746	1	TENN_PIG	O29116 sus scrofa

## ALIGNMENTS

RESULT	ID	EOST_ACTEO	STANDARD	PRT	199 AA.
AC	P81439				
DT	15-DEC-1998	(Rel. 37, Created)			
DT	15-DEC-1998	(Rel. 37, Last sequence update)			
DT	15-DEC-1998	(Rel. 37, Last annotation update)			
DE	EQUISTATIN.				
OS	Actinia equina (European sea anemone).				
OC	Eukaryota; Metazoa; Cnidaria; Anthozoa; Zantharia; Actiniaria;				
OC	Nynanthese; Actiniidae; Actinia.				
OX	NCBI_TaxID=6106;				
RN	[1]				
RP	SEQUENCE.				
RP	MEDLINE=97298102; PubMed=9153250;				
RX	Leuarcic B., Ritonja A., Strukej B., Turk B., Turk V.,				
RA	"Equistatin, a new inhibitor of cysteine proteinases from Actinia				
RT	equina, is structurally related to thyroglobulin type-1 domain."				
RL	J. Biol. Chem. 272:13899-13903(1997).				
RN	[2]				
RP	REVISIONS.				
RA	Lenarcic B., Ritonja A., Strukej B., Turk B., Turk V.,				
RL	J. Biol. Chem. 273:12682-12682(1998).				
CC	-1- FUNCTION: POTENT INHIBITOR OF PAIN-LIKE CYSTEINE PROTEINASES				
CC	AND HAS A PI OF 4.7.				
CC	-1- SIMILARITY: CONTAINS 3 THYROGLOBULIN TYPE-1 DOMAINS.				
CC	InterPro: IPR000716; Thyroglobulin_1.				
DR	Pfam: PF00086; thyroglobulin_1; 3.				
DR	SMART: SM00211; TY; 3.				
DR	PROSITE: PS00484; THYROGLOBULIN_1; 2.				
KW	Thiol protease inhibitor; Repeat.				
FT	DOMAIN 18 66				
FT	DOMAIN 86 134				
FT	DOMAIN 155 199				
FT	VARIANT 3 3				
FT	VARIANT 42 42				
FT	VARIANT 130 130				
FT	VARIANT 137 137				
FT	VARIANT 143 144				
FT	VARIANT 156 156				
FT	VARIANT 184 184				
FT	VARIANT 184 184				
SO	SEQUENCE 199 AA; 21755 MW; 87E44735F8171BRC CRC64;				

Query Match 84.6%; Score 1089; DB 1; Length 199;  
Best Local Similarity 95.5%; Pred. No. 2.76-90;  
Matches 190; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY	33	SUTKCOQLQASNSGLIGTYVPCKETGEFEFKOCWGSTGYCWCVDDEGKELGTRKINS	92
DB	1	SLSKCOQLQASNSGLIGTYVPCKETGEFEFKOCWGSTGYCWCVDDEGKELGTRKINS	60
QY	93	PCSSRRKALTLTLCOMMGAITIVVPGMCGPPSCADGSDFEVQCCASNGECYCVDRKREL	152
DB	61	PCSSRRKALTLTLCOMMGAITIVVPGMCGPPSCADGSDFEVQCCASNGECYCVDRKREL	120



Query Match Similarity 23.0% Score 295.5 DB 1 Length 2766;  
Best Local Similarity 30.3% Pred. No. 2e-18; Mismatches 83; Indels 61; Gaps  
Matches 76; Conservative 31;

QY 16 VMTVFIAFCAITSTASLTFRCK-----OLQASANSGLGIYVPCKETGFEERKQ 66  
:::| : : : : : |  
Db 4 LVLVASTLLSSVCLVAANFFEXYDAQFLRCELREKRAFLKAQEYVPCSDSGSFQYQ 63  
| : : : : : |

OY 67 CNGSTGYCWCVDDEKELIGKINGSPPCCSRKAALTLICM-MQATL---VNPGWCGP 125  
| ||||| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
Db 64 QONDGSCWCVCSDREVPGRQLGRP-----TVCLSPCOLHKORILLGSLINSTDALYL 118  
| : : : : : |

OY 122 PSCKADSEDEVOCASCANGECYCVDKKGELETRDQGRPT-CERHLSECE-EARIKAS 179  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
Db 119 PQCDGSGNAPVQCPLDLRVQCWCVDTEGMEYVGPRGRTCPR--SCEIRNRLLHG 175  
| : : : : : |

OY 180 NSLRVEMFPECLEEGSGYNPOC-----MPS- 205  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
Db 176 VGDRSP---PQCTAAGEPMPOCFKVNTDMTMIIDLHNRPDAFYTFESSRGRRPEV 233  
| : : : : : |

OY 206 TGYCMWCDEGG 216  
:|:|:|:|:|:  
Db 233 SGXYCYACDSQG 243

RESULT 3  
THYG\_BOVIN STANDARD; PRT; 2769 AA.

ID THYG\_BOVIN AC P01267; O18976; Q95478; Q28186;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 02, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE THYROGLOBULIN PRECURSOR.  
CN TC.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mamalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OX Bovidae; Bovinae; Bos.  
NCBI\_Taxid=9913;  
[1]  
RN NCBII\_Taxid=9913;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85296288; PubMed=3855243;  
RA Mercken L., Simons M.-J., Swillens S., Vassart M., Vassart G.;  
RT "Primary structure of bovine thyroglobulin deduced from the sequence  
of its 8,431-base complementary DNA.";  
RL Nature 316:647-651(1985).  
[2]  
RN NCBII\_Taxid=9913;  
RP SEQUENCE OF 1-930 FROM N.A.  
RX MEDLINE=85127025; PubMed=3855750;  
RA Mercken L., Simons M.-J., de Martynhoff G., Swillens S., Vassart G.;  
RT "Presence of hormonogenic and repellitive domains in the first 930  
amino acids of bovine thyroglobulin as deduced from the cDNA  
sequence.";  
RL Eur. J. Biochem. 147:59-64(1985).  
[3]  
RN NCBII\_Taxid=9913;  
RP SEQUENCE OF 1-22 FROM N.A.  
RX MEDLINE=87190432; PubMed=3032624;  
RA de Martynhoff G., Pohl V., Mercken L., van Ommen G.-J., Vassart G.;  
RT "Structural organization of the bovine thyroglobulin gene and of its  
5'-flanking region.";  
RL Eur. J. Biochem. 164:591-599(1987).  
[4]  
RN NCBII\_Taxid=9913;  
RP SEQUENCE OF 1002-1209 FROM N.A.  
RX MEDLINE=88062712; PubMed=3681978;  
RA Parma J., Christophe D., Pohl V., Vassart G.;  
RT "Structural organization of the 5' region of the thyroglobulin gene.  
Evidence for intron loss and 'exonization' during evolution.";  
RJ Mol. Biol. 195:769-779(1987).  
CC -I- FUNCTION: PRECURSOR OF THE IODINATED THYROID HORMONES THYROXINE  
(T4) AND TRIIODOTHYRONINE (T3).  
CC -I- SUBUNIT: HOMODIMER.

	CC	-I-	SUBCELLULAR LOCATION:	SECRETED.
	CC	-I-	TISSUE SPECIFICITY:	THYROID GLAND SPECIFIC.
	CC	-I-	PTM:	SULFATED (BY SIMILARITY).
	CC	-I-	SIMILARITY:	BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
	CC	-I-	SIMILARITY:	CONTAINS 11 THYROGLOBULIN TYPE-1 DOMAINS.
	CC			-----
	CC			This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isdb.ch/announce/ or send an email to licenseseis@isdb.ch).
	CC			-----
DR	EMBL;	X02815;	CMA26584.1; .	
DR	EMBL;	X02155;	CMA26090.1; .	
DR	EMBL;	X05380;	CNA28971.1; ALT_SEQ.	
DR	EMBL;	X06071;	CAA29457.1; .	
DR	EMBL;	X06072;	CMA29457.1; JOINED.	
DR	EMBL;	X06073;	CMA29457.1; JOINED.	
DR	EMBL;	X06074;	CMA29457.1; JOINED.	
DR	EMBL;	X06075;	CMA29457.1; JOINED.	
DR	PIR;	A01533;	UIBO.	
DR	HSSP;	P21836;	IMA#.	
DR	InterPro;	IPR002018;	Carboxylesterase_B.	
DR	InterPro;	IPR000716;	Thyroglobulin_1.	
DR	pfam;	PF00135;	Coesterase_2.	
DR	pfam;	PF00086;	thyroglobulin_1; 8.	
DR	SMART;	SMO0211;	TY; 10.	
DR	PROSITE;	PS00484;	THYROGLOBULIN_1; 9.	
DR	PROSITE;	PS00941;	CARBOXYLESTERASE_B_2; 1.	
KW	Glycoprotein;	Repeat;	Thyroid hormone; Iodination; Sulfation; Signal.	
FT	SIGNAL	1..19		
FT	CHAIN	20..2769	THYROGLOBULIN.	
FT	DOMAIN	31..92	THYROGLOBULIN TYPE IA 1.	
FT	DOMAIN	93..160	THYROGLOBULIN TYPE IA 2.	
FT	DOMAIN	161..297	THYROGLOBULIN TYPE IA 3.	
FT	DOMAIN	298..358	THYROGLOBULIN TYPE IA 4.	
FT	DOMAIN	604..657	THYROGLOBULIN TYPE IA 5.	
FT	DOMAIN	661..725	THYROGLOBULIN TYPE IA 6.	
FT	DOMAIN	726..921	THYROGLOBULIN TYPE IA 7.	
FT	DOMAIN	922..1073	THYROGLOBULIN TYPE IA 8.	
FT	DOMAIN	1074..1145	THYROGLOBULIN TYPE IB 1.	
FT	DOMAIN	1146..1210	THYROGLOBULIN TYPE IA 9.	
FT	REPEAT	1458..1471	TYPE II.	
FT	REPEAT	1472..1488	TYPE II.	
FT	REPEAT	1489..1505	TYPE II.	
FT	DOMAIN	1513..1567	THYROGLOBULIN TYPE IB 2.	
FT	REPEAT	1605..1725	TYPE IIIA.	
FT	REPEAT	1726..1893	TYPE IIIB.	
FT	REPEAT	1894..1996	TYPE IIIC.	
FT	REPEAT	1997..2130	TYPE IIID.	
FT	REPEAT	2131..2188	TYPE IIIE.	
FT	MOD_RES	24..24	IODINATION (IN T4).	
FT	MOD_RES	24..24	SOLENATION (POTENTIAL).	
FT	MOD_RES	2574..2574	IODINATION (IN T4).	
FT	MOD_RES	2588..2588	IODINATION (IN T4).	
FT	MOD_RES	2767..2767	IODINATION (IN T3).	
FT	CARBOHYD	110..110	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	198..198	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	483..483	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	495..495	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	747..747	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	853..853	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	947..947	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1140..1140	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1365..1365	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1776..1776	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1870..1870	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2014..2014	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2123..2123	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2251..2251	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2296..2296	N-LINKED (GLCNAC. .)	(POTENTIAL).

FT CONFLICT 1206 1206 S -> R (IN REF. 4).  
 SQ SEQUENCE 2769 AA; 303218 MW; 1C7F227E9101DE2A CRC64;  
 Query Match 22.9%; Score 295; DB 1; Length 2769;  
 Best Local Similarity 33.5%; Pred. No. 2,2e-18;  
 Matches 69; Conservative 21; Mismatches 64; Indels 52; Gaps 8;

QY 52 YPQCKEPEFEKOCWSTGTCWCVDDEGKEILGTRISPPDCSRKALTLQMQMAY 111  
 DB 48 YPQCAEDSFTQVCCGADGASCWCVDADGREGVSGRQGRP-----ACLSFCQLOKQO 102  
 QY 112 I-----VNPYGCWCPSCKADSPEDVCCASNGECYCKKKELEGTRQGRPT-CER 165  
 DB 103 ILLSYINSTATSYLPQCGDSDGYSVPCDRLRRQCCWCVDAGMEYGYTRQGRPARCPR 162  
 QY 166 HLESECF-EARIRAHNSLSLRYEMFVECELEDSYNPVOC----- 202  
 DB 163 ---SCEIRNRRLHGVGDSP---PQCSPDGAFRPPVCKLWNTDMKIFDLVHSSRFPD 216  
 QY 203 -----NPS-TGYCWCVDDEG 216  
 DB 217 APTFSSFRSFRPEVSGCYCADSG 242

RESULT 4  
 THYR\_RAT STANDARD; PRT; 2768 AA.  
 AC P06882; Q9JMG4; Q9JKY6;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE THYROGLOBULIN PRECURSOR.  
 GN TG.  
 OS Rattus norvegicus (rat).  
 OC Eukaryota; Metazoa; Chordata; Placenta; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NIHSTAR IMAMICHI;  
 RA Hisahuma A.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISCHER 344;  
 RA Ding M., Jung C.-C., Cheng J.-M., Miyamoto T., Furudate S.I.,  
 RA Agui T.;  
 RT "A missense mutation in the thyroglobulin gene causes hypothyroidism  
 RT and dwarfism not associated with goiter in the MIC-rdw rat";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-213 FROM N.A.  
 RC STRAIN=FISCHER; TISSUE=Thymocytes;  
 RA MEDLINE=90220642; PubMed=2325666;  
 RA Graves P.N., Davies T.F.;  
 RT "A second thyroglobulin messenger RNA species (Tg-2) in rat  
 RT thyrocytes";  
 RL Mol. Endocrinol. 4:155-161(1990).  
 RN [4]  
 RP SEQUENCE OF 1-59 FROM N.A.  
 RC MEDLINE=86094383; PubMed=3455768;  
 RA Musti A.M., Avvedimento V.E., Polistina C., Ursini V.M., Obici S.,  
 RA Nitsch L., Cocozza S., di Lauro R.;  
 RT "The complete structure of the rat thyroglobulin gene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:323-327(1986).  
 RN [5]  
 RP SEQUENCE OF 1802-2768 FROM N.A.  
 RC MEDLINE=85154044; PubMed=3838512;  
 RA di Lauro R., Obici S., Condiliffe D., Ursini V.M., Musti A.M.,  
 RA Moscatelli C., Avvedimento V.E.;  
 RT "The sequence of 967 amino acids at the carboxyl-end of rat  
 RT thyroglobulin. Location and surroundings of two tyroxine-forming

RT sites."; Biochem. 148:7-11(1985).  
 RL Eur. J. Biochem. 148:7-11(1985).  
 CC -1 FUNCTION: PRECURSOR OF THE IODINATED THYROID HORMONES THYROXINE  
 CC (T4) AND TRIIODOTHYRONINE (T3).  
 CC -1 SUBUNIT: HOMODIMER.  
 CC -1 SURCELLULAR LOCATION: SECRETED.  
 CC -1 TISSUE SPECIFICITY: THYROID GLAND SPECIFIC.  
 CC -1 PTM: SUTURED (BY SIMILARITY).  
 CC -1 MISCELLANEOUS: IT IS NOT CERTAIN WHETHER THIS THYROGLOBULIN PLAYS  
 CC ANY ROLE IN THE FORMATION OF TRIODOTHYRONINE.  
 CC -1 SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.  
 CC -1 SIMILARITY: CONTAINS 11 THYROGLOBULIN TYPE-1 DOMAINS.  
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION  
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -  
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS  
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY  
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL  
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb.ch/annouce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB035201; BAA6132.1; -.  
 DR EMBL: AF221622; AAF34909.1; -.  
 DR EMBL: M35965; AAA42089.1; ALT\_TERM.  
 DR EMBL: M12559; AAA50379.1; JOINED.  
 DR EMBL: M12558; AAA50379.1; JOINED.  
 DR EMBL: X02318; CAA26183.1; -.  
 DR PIR: A22016; UIRT.  
 DR HSSP: P21836; IMAH.  
 DR InterPro: IPR002018; Carboxylesterase-B.  
 DR InterPro: IPR007116; Thyroglobulin\_1.  
 DR Pfam: PF00135; Coesterase\_2.  
 DR Pfam: PF00086; thyroglobulin\_1; 9.  
 DR SMART: SM00211; TY; 10.  
 DR PROSITE: PS00484; THYROGLOBULIN\_1; 9.  
 DR PROSITE: PS00941; CARBOXYLESTERASE-B\_2; 1.  
 DR GlycoProtex: Repeat; Thyroid hormone; Iodination; Sulfation; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766  
 FT CARBOHYD 199 111  
 FT CARBOHYD 199 199  
 FT CARBOHYD 484 484  
 FT CARBOHYD 496 496  
 FT CARBOHYD 545 545  
 FT CARBOHYD 748 748  
 FT CARBOHYD 817 817  
 FT CARBOHYD 948 948  
 FT CARBOHYD 1017 1017

FT SIGNAL 1 20  
 FT CHAIN 21 2768  
 FT DOMAIN 32 93  
 FT DOMAIN 94 161  
 FT DOMAIN 162 298  
 FT DOMAIN 289 359  
 FT DOMAIN 605 658  
 FT DOMAIN 727 922  
 FT DOMAIN 923 1074  
 FT DOMAIN 1075 1146  
 FT DOMAIN 1147 1211  
 FT REPEAT 1455 1468  
 FT REPEAT 1469 1485  
 FT REPEAT 1486 1502  
 FT DOMAIN 1510 1564  
 FT REPEAT 1602 1722  
 FT REPEAT 1723 1891  
 FT REPEAT 1892 1994  
 FT REPEAT 1995 2127  
 FT REPEAT 2128 2185  
 FT DISULFID 2265 2282  
 FT MOD\_RES 25 25  
 FT MOD\_RES 25 25  
 FT MOD\_RES 2574 2574  
 FT MOD\_RES 2588 2588  
 FT MOD\_RES 2766 2766

```

FT CARBOHYD 1141 1141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1349 1349 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1365 1365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1715 1715 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1773 1773 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1866 1866 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1937 1937 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2012 2012 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2122 2122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2251 2251 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2296 2296 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2445 2445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2583 2583 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 44 44 L -> V (IN REF. 4).
FT CONFLICT 678 678 A -> V (IN REF. 2).
FT CONFLICT 1492 1492 C -> F (IN REF. 2).
FT CONFLICT 1732 1732 CC -> KS (IN REF. 2).
FT CONFLICT 1914 1914 L -> F (IN REF. 5).
FT CONFLICT 2043 2043 R -> A (IN REF. 5).
FT CONFLICT 2081 2081 Q -> K (IN REF. 5).
FT CONFLICT 2126 2126 A -> V (IN REF. 5).
FT CONFLICT 2153 2153 R -> K (IN REF. 5).
FT CONFLICT 2169 2169 S -> N (IN REF. 5).
FT CONFLICT 2611 2611 M -> I (IN REF. 5).
FT CONFLICT 2658 2658 Q -> H (IN REF. 5).
SQ SEQUENCE 2768 AA; 304642 MW; 290DD6943F23F3D CRC64;

Query Match 22.9%; Score 294.5; DB 1; Length 2768;
Best Local Similarity 30.3%; Pred. No. 2,4e-18;
Matches 76; Conservative 32; Mismatches 82; Indels 61; Gaps 10;

QY 16 VAIWVLFICAITSTESATKCO-----QLQASANSGLIGTYPPCKETGEFEENEKQ 66
DB 4 LVIVSTLLSVCLVAIAINFEQVDAQPLRCELQREKAFKODEVYFPQCSSEGSFQTVQ 63
QY 67 CMGSTGYCWCVDGKELGTIRKSPDCSRKALTLCOM-MQAIT---VVPWPCGP 121
DB 64 CONDGSCWCVDGSEVPGRQLGRP-----TACLSFCQLHMQRLTLLSYINSTALVL 118
QY 122 PSCKADGSEFDEVOCCASNGECYDKKKKELEGTROGGRPT-CEPHLSDEE-EARKKALS 179
DB 119 PQODSGNAPVQCDLQOVQWCVDTESEMEYGTROGGRPTCRP---SCELSRRLHLHG 175
QY 180 NSLRVEMFVPECLDEGSYNPVOC-----WPS- 205
DB 176 VG---DKSPQCQDADGGEFMPVQCKFVNTDMIFDLIHYNRFPDAVFSAFRNRPV 232
QY 206 TGYCWCVDGSG 216
DB 233 SGYCYCADSOG 243

RESULT 5
THYG_HUMAN STANDARD; PRT; 2768 AA.
AC P01266; O15593; O43899; O15274; Q9NVR1; Q9NVR2; Q9UWZ0; Q9UNY3;
DT 21-JUL-1986 (Rel. 01, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE THYROGLOBULIN PRECURSOR.
GN TG.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8724630; PubMed=359559;
RA Malthiery Y., Lissitzky S.;
RT "Primary structure of human thyroglobulin deduced from the sequence
RT of its 8448-base complementary DNA.";
RL Eur. J. Biochem. 165:491-498(1987).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=97329813; PubMed=9186272;
RA van de Graaf S.A.R., Pauw E., de Vijlder J.J.M., Ris-Stalpers C.;
RT "The revised 8307 base pair coding sequence of human thyroglobulin
RT transiently expressed in eukaryotic cells.";
RL Eur. J. Endocrinol. 136:508-515(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Blechschmidt K., Kalaydjieva L., Goodman R., Gresham D., Baas F.,
RA de Jonge R., Schlihaber M.B., Menzel U., Dette M.D., Baumgart C.,
RA Jahn N., Rosenthal A.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-730 FROM N.A.
RX MEDLINE=85127024; PubMed=3971976;
RA Malthiery Y., Lissitzky S.;
RT "Sequence of the 5'-end quarter of the human-thyroglobulin messenger
RT ribonucleic acid and of its deduced amino-acid sequence.";
RL Eur. J. Biochem. 147:53-58(1985).
RN [5]
RP SEQUENCE OF 1-415; 640-737 AND 880-1000 FROM N.A.
RX MEDLINE=88062712; PubMed=3683978;
RA Parma J., Christophe D., Pohl V., Vassart G.;
RT "Structural organization of the 5' region of the thyroglobulin gene.
RT Evidence for intron loss and 'exonization' during evolution.";
RL J. Mol. Biol. 196:769-779(1987).
RN [6]
RP SEQUENCE OF 1-25 FROM N.A.
RX MEDLINE=85269632; PubMed=2991855;
RA Christophe D., Cabrer B., Bacolla A., Targovnik H.M., Pohl V.,
RA Vassart G.;
RT "An unusually long poly(purine)-poly(pyrimidine) sequence is located
RT upstream from the human thyroglobulin gene.";
RL Nucleic Acids Res. 13:5127-5144(1985).
RN [7]
RP SEQUENCE OF 1002-1566 FROM N.A.
RX PubMed=11124863;
RA Moya C.M., Mendive F.M., Rivolta C.M., Vassart G., Targovnik H.M.;
RT "Genomic organization of the 5' region of the human thyroglobulin
RT gene.";
RL Eur. J. Endocrinol. 143:789-798(2000).
RN [8]
RP SEQUENCE OF 1645-2768 FROM N.A.
RX MEDLINE=99452200; PubMed=10524569;
RA Mendive F.M., Rivolta C.M., Vassart G., Targovnik H.M.;
RT "Genomic organization of the 3' region of the human thyroglobulin
RT gene.";
RL Thyroid 9:903-912(1999).
RN [9]
RP PRESENCE OF A 11TH THYROGLOBULIN TYPE-I REPEAT.
RX MEDLINE=96390872; PubMed=8797845;
RA Molina F., Bouanani M., Pau B., Granier C.;
RT "Characterization of the type-1 repeat from thyroglobulin, a
RT cysteine-rich module found in proteins from different families.";
RL Eur. J. Biochem. 240:125-133(1996).
RN [10]
RP IODINATION SITES.
RX MEDLINE=89340430; PubMed=2760035;
RA Lamas L., Anderson P.C., Fox J.W., Dunn J.T.;
RT "Consensus sequences for early iodination and hormonogenesis in human
RT thyroglobulin.";
RL J. Biol. Chem. 264:13541-13545(1989).
RN [11]
RP SULFATION.
RX PubMed=10448091;
RA Nlend M.-C., Cauvi D., Venot N., Chabaud O.;
RT "Sulfated tyrosines of thyroglobulin are involved in thyroid hormone
RT synthesis.";
RL Biochem. Biophys. Res. Commun. 262:193-197(1999).
RN [12]
RP VARIANT GOITER HIS-870.

```



CC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF N-TERMINUS.  
 RN MEDLINE=98406162; PubMed=9733643;  
 RA Kohfeldt E., Sasaki T., Goehring W., Timpi R.;  
 RT "Nidogen-2: a new basement membrane protein with diverse binding  
 RT properties";  
 RL J. Mol. Biol. 282:99-109(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cancellous bone;  
 RA Ohno I., Hashimoto J., Takaoka K., Ochi T., Okubo K., Matsubara K.;  
 RT "The cloning and characterization of a cDNA for the novel bone matrix  
 RT protein: osteonodogen";  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Ohno I., Okubo K., Matsubara K.;  
 RT "Human osteonodogen gene: Intron-exon junctions and chromosomal  
 RT localization";  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CELL ADHESION GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED  
 CC IN BASEMENT MEMBRANES. BINDS TO COLLAGENS I AND IV, TO PERLECAN  
 CC AND TO LAMININ 1. DOES NOT BIND FIBRIN. IT PROBABLY HAS A ROLE  
 CC IN CELL-EXTRACELLULAR MATRIX INTERACTIONS.  
 CC -1- TISSUE SPECIFICITY: HEART, PLACENTA AND BONE. LESS IN PANCREAS,  
 CC KIDNEY AND SKELETAL MUSCLE.  
 CC -1- PTM: HIGHLY N- AND O-GLYCOSYLATED.  
 CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 THYROGLOBULIN TYPE-1 DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR WYTD DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isp.slb.ch/announce/>  
 CC or send an email to [license@isp.slb.ch](mailto:license@isp.slb.ch))  
 CC -----  
 DR EMBL: AJ23500; CAA11418.1; -  
 DR EMBL: D86423; BAA13087.1; -  
 DR EMBL: AB009799; BAA24112.1; -  
 DR EMBL: AB009778; BAA24112.1; JOINED.  
 DR EMBL: AB009779; BAA24112.1; JOINED.  
 DR EMBL: AB009780; BAA24112.1; JOINED.  
 DR EMBL: AB009781; BAA24112.1; JOINED.  
 DR EMBL: AB009782; BAA24112.1; JOINED.  
 DR EMBL: AB009783; BAA24112.1; JOINED.  
 DR EMBL: AB009784; BAA24112.1; JOINED.  
 DR EMBL: AB009785; BAA24112.1; JOINED.  
 DR EMBL: AB009786; BAA24112.1; JOINED.  
 DR EMBL: AB009787; BAA24112.1; JOINED.  
 DR EMBL: AB009788; BAA24112.1; JOINED.  
 DR EMBL: AB009789; BAA24112.1; JOINED.  
 DR EMBL: AB009790; BAA24112.1; JOINED.  
 DR EMBL: AB009791; BAA24112.1; JOINED.  
 DR EMBL: AB009792; BAA24112.1; JOINED.  
 DR EMBL: AB009793; BAA24112.1; JOINED.  
 DR EMBL: AB009794; BAA24112.1; JOINED.  
 DR EMBL: AB009795; BAA24112.1; JOINED.  
 DR EMBL: AB009796; BAA24112.1; JOINED.  
 DR EMBL: AB009797; BAA24112.1; JOINED.  
 DR EMBL: AB009798; BAA24112.1; JOINED.  
 DR MIM: 605339;  
 DR InterPro: IPR000152; ASX\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR000033; Ldl\_receptor\_rep.  
 DR InterPro: IPR003886; Nidogen\_ext.  
 DR InterPro: IPR000716; Thyroglobulin\_1.  
 DR Pfam: PF00008; EGF\_5.

DR Pfam: PF00058; Ldl\_recept\_b; 4.  
 DR Pfam: PF00086; thyroglobulin\_1; 2.  
 DR SMART: SM00179; EGF\_Ca; 2.  
 DR SMART: SM00001; EGF\_Like; 3.  
 DR SMART: SM00135; LY; 4.  
 DR SMART: SM00539; NIDO; 1.  
 DR SMART: SM00211; TY; 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 3.  
 DR PROSITE: PS00484; THYROGLOBULIN\_1; 2.  
 DR PROSITE: PS01186; EGF\_2; 4.  
 DR PROSITE: PS01187; EGF\_Ca; 2.  
 DR Basal membrane: Extracellular matrix; Glycoprotein; Signal;  
 KW Calcium-binding; Repeat; EGF-like domain; Cell adhesion.  
 FT SIGNAL 1 30  
 FT CHAIN 31 1375  
 FT DOMAIN 484 524  
 FT DOMAIN 759 800  
 FT DOMAIN 801 843  
 FT DOMAIN 848 891  
 FT DOMAIN 892 930  
 FT DOMAIN 936 1006  
 FT DOMAIN 1015 1085  
 FT DOMAIN 1157 1200  
 FT DOMAIN 1201 1243  
 FT DOMAIN 1244 1288  
 FT DOMAIN 1289 1328  
 FT DOMAIN 1329 1373  
 FT DISULFID 763 776  
 FT DISULFID 770 786  
 FT DISULFID 788 799  
 FT DISULFID 805 818  
 FT DISULFID 812 827  
 FT DISULFID 829 842  
 FT DISULFID 852 867  
 FT DISULFID 859 877  
 FT DISULFID 879 890  
 FT DISULFID 896 907  
 FT DISULFID 901 916  
 FT DISULFID 918 929  
 FT CARBOHYD 417 417  
 FT CARBOHYD 658 658  
 FT CARBOHYD 693 693  
 FT CARBOHYD 703 703  
 FT CARBOHYD 1124 1124  
 FT CONFLICT 44 44  
 FT CONFLICT 54 172  
 FT CONFLICT 453 453  
 FT SEQUENCE 1375 AA; 151394 MW; 414299244205FBC CRC64;  
 SQ  
 Query Match 19.18; Score 245.5; DB 1; Length 1375;  
 Best Local Similarity 30.88; Pled. No. 2.9e-14;  
 Matches 62; Conservative 23; Mismatches 55; Indels 61; Gaps 7;  
 OY 27 ITSTFASLTTCQOIQ--ASANSGLIGT--VVPCKTGTGFEKQCGSTGYCMVDGDK 82  
 DB 930 IDSTSSLTPEOQQRHQAQYANPGARFIHQCDQGNFLPLQCHGSTGFCVCDPBGH 969  
 OY 83 ELTKIKINGSPDCSKRRKALTLQCMQALIVNPPGMCQPPSCADKSGPEVQCCASNEC 142  
 DB 990 EVPGTQ-----TPGSTPHC----- 1005  
 OY 143 YCVDRKGRLEGTROQGPT--CERNHLSCEBEARIKASHSLRYEMFVEPCLEDGSSYNPQ 201  
 DB 1006 -----GSPPT--QRPPTICR-----WRENLEHYGGTTPDDQYVPCODLGHFTLQ 1053  
 OY 202 CWPSTGYCMVCDGEGVAVPGS 222

FT	DOMAIN	109	251	ABSENT IN TRANSFERRINS.
FT	DISULFID	29	64	BY SIMILARITY.

FT	DISULFID	322	335	BY SIMILARITY.
----	----------	-----	-----	----------------

FT	DISULFID	388	402	BY SIMILARITY.
----	----------	-----	-----	----------------

FT	DISORDER	503	318	BY SIMILARITY.
EM	DISORDER	553	839	BY SIMILARITY.

ET DISULFID 607 685 BY SIMILARITY.

FT DISULFID 652 668 BY SIMILARITY.

SEQUENCE 844 AA; 93089 MW; 162A/65AD02C3C5E CRC64;

Query Match 19 0%: Score 244: DB 1: Length 844:

Matches 57; Conservative 25; Mismatches 47; Indels 68; Gaps 5;

29 STEASLTKC-QQLQASANSGLIGTYVPQCKETGEFFEEKQCWGSTGYCWCVDEDEGKEILGT 87

DD 102 35NKULYNLEAKRYDALANMIGHI 1PQCDENGINI YPQCHGSI GHWCVNAMGENLSGT 101

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

167 -----OTBATEPHEI.PKTI.KEROVAIGDEKVI.GREYBOSDEKNGVEBOEHGS 216

QY 206 TGYCWCVDDEGGVKVPGS 222

Db 217 TGYSWCVNAIGEEIAGT 233 .

DESTROY 8

-----  
ID NID2 MOUSE STANDARD; PRT; 1403 AA.

DT 20-AUG-2001 (Rel. 40, Created)

DI 20-AUG-2001 (Rel. 40, Last annotation update)

Mus musculus (Mouse).

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

KN [L]  
PB SENTENCE FROM N A

RA Kimura N.; Toyoshima T.; Kojima T.; Shimane M.;

Rt homology to entactin/nidogen.";

CC -1- FUNCTION: CELL ADHESION GLYCOPROTEIN. MIGHT BE INVOLVED IN  
CC OSTEOBLAST DIFFERENTIATION. THROBOPARIV HAS A ROLE IN CELL-

CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.

CC -I- SIMILARITY; CONTAINS 5 EGF-LIKE DOMAINS.

-!- SIMILARITY; CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.

between the Swiss Institute of Bioinformatics and the EMBL outstation.

CC use by non-profit institutions as long as its content is in no way

1



```

Db 1038 -----BPT-QRPTVGER-----WRSLSLEHGGGPRDDQYVPCDDILGHRTPLQJCHKXSD 1087
Oy 208 YCMCVDGKVKVPGSDVR-FKRPCT 231
Db 1088 FCWCVDKDKGRELQGTGRSPGTGRPAC 1112

RESULT 9
HG2A_MOUSE STANDARD: PRT, 279 AA.
AC P04441;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN, GAMMA CHAIN (MHC CLASS II(A)
DE ASSOCIATED INVARIANT CHAIN).
CD 74 OR 11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
RA RT "Complete sequence of the murine invariant chain (Ii) gene."
RL Nucleic Acids Res. 17:447-448(1989).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=AKR;
RX MEDLINE=67275861; PubMed=3038530;
RA Koch N., Lauer W., Habicht J., Dobberstein B.;
RT "Primary structure of the gene for the murine Ia antigen-associated
RT invariant chains (Ii). An alternatively spliced exon encodes a
RT cysteine-rich domain highly homologous to a repetitive sequence of
RT thyroglobulin."
RL EMBO J. 6:1677-1683(1987).
[3]
RA RT SEQUENCE OF 1-58 FROM N.A.
RP Stone J., Perry R., Todd J.A., McDevitt H.O.;
RT "Nucleotide sequences of the murine Ia-associated invariant chain (Ii)
RT and I-E (H-2S, Beta) chain expressible cDNA clones."
RL submitted (MAR-1988) to the EMBL/Genbank/DBJ databases.
[4]
RA RT SEQUENCE OF 47-191 AND 256-279 FROM N.A.
RP MEDLINE=64207946; PubMed=6327293;
RA Singer P.A., Lauer W., Dembic Z., Mayer W.E., Lipp J., Koch N.,
RA Hammerling G., Klein J., Dobberstein B.;
RT "Structure of the murine Ia-associated invariant (Ii) chain as
RT deduced from a cDNA clone."
RL EMBO J. 3:873-877(1984).
[5]
RA RT FUNCTION: PLAYS A CRITICAL ROLE IN MHC CLASS II ANTIGEN PROCESSING
CC BY STABILIZING PEPTIDE-FREE CLASS II ALPHA/BETA HETERODIMERS IN A
CC COMPLEX SOON AFTER THEIR SYNTHESIS AND DIRECTING TRANSPORT OF THE
CC COMPLEX FROM THE ENDOPLASMIC RETICULUM TO COMPARTMENTS WHERE
CC PEPTIDE LOADING OF CLASS II TAKES PLACE.
CC -1- SUBUNIT: NONAER COMPOSED OF THREE ALPHA/BETA/GAMMA HETEROTRIMERS
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/sib.ch).
CC or send an email to license@sib-sib.ch).
DR EMBL: X00496; CAA25191.1; -
DR EMBL: X07129; CAA30141.1; -
DR

```

DR EMBL: X05428; CAA29010.1; -;  
 DR EMBL: X05429; CAA29012.1; -;  
 DR EMBL: X05430; CAB37297.1; -;  
 DR EMBL: X13414; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: A02244; HLHMSG.  
 DR PIR: A27866; A27866.  
 DR PIR: B27866; B27866.  
 DR PIR: S03099; S03099.  
 DR PIR: S09284; S09284.  
 DR HSSP: P04233; 1A6A.  
 DR MGD: MGI:96534; 11.  
 DR InterPro: IPR000716; Thyroglobulin\_1.  
 DR Pfam: PF00086; thyroglobulin\_1; 1.  
 DR SMART: SMO0211; TY; 1.  
 DR PROSITE: PS00484; THYROGLOBULIN\_1; 1.  
 DR Glycoprotein: MHC; Proteoglycan; Transmembrane; Alternative splicing;  
 KW Signal-anchor.  
 FT DOMAIN 1 29 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 30 55 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT POTENTIAL.  
 FT DOMAIN 56 279 EXTRACELLULAR (POTENTIAL).  
 FT CARBOHYD 113 113 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 119 119 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 265 265 O-LINKED (GLYCOSAMINOGLYCAN).  
 FT DOMAIN 211 254 THYROGLOBULIN TYPE I.  
 FT VARSPLIC 192 255 MISSING (IN SHORT ISOFORM).  
 FT CONFLICT 10 10 N -> M (IN REF. 1).  
 FT CONFLICT 113 113 N -> Q (IN REF. 1).  
 FT CONFLICT 229 231 STG -> RHC (IN REF. 2).  
 FT CONFLICT 231 231  
 SQ SEQUENCE 279 AA; 31557 MW; 300A85014F170792 CRC64;

Query Match 12.1%; Score 156; DB 1; Length 279;  
 Best Local Similarity 40.7%; Pred. No. 5, 6e-07;  
 Matches 35; Conservative 11; Mismatches 32; Indels 8; Gaps 2;

QY 19 WYFLACAT-----STAS---LTKCOQLQASNSGLITTYPOCKEGEREKCQKMS 70  
 DB 170 WLFEMSKNSLEKKRPTAPKRVLTCKOEVSHPVAYPGAFPKCDGENTYLPQCHGS 229  
 QY 71 TGYCWCVDGDEKKEILGTRKIRSPDSCS 96  
 DB 230 TGYCWCVFNGTEVPHTRKSRHNCSS 255

RESULT 10  
 NIDO\_MOUSE STANDARD; PRT; 1245 AA.  
 AC P10493;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NIDOGEN PRECURSOR (ENTACTIN).  
 GN NID OR NIDI OR ENT.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40.  
 RX MEDLINE=89079780; PubMed=3264556;  
 RA Durkin M.E., Chakravarti S., Barros B.B., Liu S.H., Friedman R.L.,  
 RA Chung A.E.,  
 RT "Amino acid sequence and domain structure of entactin. Homology with  
 RT epidermal growth factor precursor and low density lipoprotein  
 RT receptor.";  
 RL J. Cell Biol. 107:2749-2756(1988).  
 [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=89231638; PubMed=2496973;  
 RA Mann K., Deutzmann R., Aumailley M., Timpl R., Raimondi L.,  
 RA Yamada Y., Pan T.-C., Conway D., Chu M.-L.,  
 RA "Amino acid sequence of mouse nidogen, a multidomain basement

RT membrane protein with binding activity for laminin, collagen IV and  
 RT cells.";  
 RL EMBL J. 8:65-72(1989).  
 [3]  
 RP SEQUENCE OF 1-251 FROM N.A.  
 RC STRAIN-BALB/C; TISSUE=Liver;  
 RX MEDLINE=94040771; PubMed=8224873;  
 RA Durkin M.E., Liu S.H., Reing J., Chung A.E.;  
 RT "Characterization of the 5' end of the mouse Ent gene encoding the  
 RT basement membrane protein, entactin.";  
 RL Gene 132:261-266(1993).  
 [4]  
 RP SEQUENCE OF 1207-1245 FROM N.A.  
 RC STRAIN=C57BL/6J X CBA/J;  
 RX MEDLINE=95324912; PubMed=7601446;  
 RA Durkin M.E., Meyer U.M., Chung A.E.;  
 RT "Exon organization of the mouse entactin gene corresponds to the  
 RT structural domains of the polypeptide and has regional homology to  
 RT the low-density lipoprotein receptor gene.";  
 RL Genomics 26:219-228(1995).  
 [5]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE=86192477; PubMed=3084254;  
 RA Paulsson M., Deutzmann R., Dziadek M., Nowack H., Timpl R., Weber S.,  
 RA Engel J.;  
 RT "Purification and structural characterization of intact and  
 RT fragmented nidogen obtained from a tumor basement membrane.";  
 RL Eur. J. Biochem. 156:467-478(1986).  
 [6]  
 RP CARBOHYDRATE-LINKAGE SITES, AND PARTIAL SEQUENCE.  
 RX MEDLINE=93316903; PubMed=8326911;  
 RA Fujiwara S., Shinkai H., Mann K., Timpl R.;  
 RT "Structure and localization of O- and N-linked oligosaccharide chains  
 RT on basement membrane protein nidogen.";  
 RL Matrix 13:215-222(1993).  
 CC -1- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN  
 CC BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.  
 CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-  
 CC EXTRACELLULAR MATRIX INTERACTIONS.  
 CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.  
 CC -1- PTM: N- AND O-GLYCOSYLATED.  
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC  
 CC EMBL: X14194; CAA32408.1; -;  
 CC EMBL: X14480; CAA32642.1; -;  
 CC EMBL: L17324; AAAT7652.1; -;  
 CC EMBL: L17322; AAAT7652.1; JOINED.  
 CC EMBL: L17323; AAAT7652.1; JOINED.  
 CC EMBL: X83093; CAA58148.1; -;  
 CC PIR: S02730; MMSMSND.  
 CC HSSP: P07204; 2NDX.  
 CC CarBank; CCSD:37018; -;  
 CC CarBank; CCSD:37019; -;  
 CC MGD: MGI:97342; NId1.  
 CC InterPro: IPR000152; Asx\_hydroxyl.  
 CC InterPro: IPR000561; EGF-like.  
 CC InterPro: IPR001881; EGF Ca.  
 CC InterPro: IPR000033; LDL\_receptor\_rep.  
 CC InterPro: IPR003886; Nidogen\_ext.  
 CC InterPro: IPR000716; Thyroglobulin\_1.  
 CC Pfam: PF00009; EGF; 6.  
 CC Pfam: PF00058; Idl\_receptor; 3.  
 CC Pfam: PF00086; thyroglobulin\_1; 1.

[illegible]

Db	771	QRAQCIYMGSSSYTSCLPGLFSG--DGNACDVEDQCHSRHCPAFYNTPGSFTQCKP	828
Oy	149	GKELEGR-----QQGPRTCERHLSCECARPKANSLSRVE-MFVPECLDGSYNPQCM	203
Db	829	GYGDDGFCRCPMGVSXKTRCLEREHLLIAGACGADAORPTLOGMFVPCDCEYGHVYPTQCH	888
Oy	204	PTGPGYCMCVDEGSVKVPGSDV-RFRPRPC	231
Db	889	HSYGYCMCVDRGDRGLESGSRTPPGRRPFC	917
RESULT	11		
HC2A_RAT		STANDARD:	PRT: 280 AA.
AC	P10247;		
DT	01-MAR-1989 (Rel. 10, Created)		
DT	01-FEB-1991 (Rel. 17, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN, GAMMA CHAIN (HLA-DR ANTIGENS DE ASSOCIATED INVARIANT CHAIN).		
GN	CD74.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-PVG X DA; TISSUE-Spleen;		
RA	MEDLINE=69282409; PubMed=2499873;		
RT	McKnight A.J., Mason D.W., Barclay A.N.;		
RL	"Sequence of a rat MHC class II-associated invariant chain cDNA clone containing a 64 amino acid thymoglobulin-like domain."		
RL	Nucleic Acids Res. 17:3983-3984(1989).		
RN	[2]		
RP	SEQUENCE OF 1-192 AND 257-280 FROM N.A.		
RC	STRAIN-LEMIS;		
RA	MEDLINE=69098337; PubMed=3264906;		
RT	Henkes W., Syha J., Reske K.;		
RL	"Nucleotide sequence of rat invariant gamma chain cDNA clone pLR gamma 34.3."		
RL	Nucleic Acids Res. 16:11822-11822(1988).		
CC	-1- FUNCTION: PLAYS A CRITICAL ROLE IN MHC CLASS II ANTIGEN PROCESSING		
CC	BY STABILIZING PEPTIDE-FREE CLASS II ALPHA/BETA HETERODIMERS IN A		
CC	COMPLEX SOON AFTER THEIR SYNTHESIS AND DIRECTING TRANSPORT OF THE		
CC	COMPLEX FROM THE ENDOPLASMIC RETICULUM TO COMPARTMENTS WHERE		
CC	PEPTIDE LOADING OF CLASS II TAKES PLACE.		
CC	-1- SUBUNIT: NONAMER COMPOSED OF THREE ALPHA/BETA/GAMMA HETERO trimers		
CC	(BY SIMILARITY).		
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).		
CC	-1- SIMILARITY: CONTAINS 1 THYMOGLOBULIN TYPE-I DOMAIN.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on ways		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; X14254; CAA32468.1; -		
DR	EMBL; X13044; CAA31450.1; -		
DR	PIR; S02182; S02182.		
DR	PIR; S04362; S04362.		
DR	HSSP; P04233; IAG6.		
DR	InterPro; IPR000716; Thyroglobulin_1.		
DR	Pfam; PF00086; thyroglobulin_1; 1.		
DR	SMART; SM00211; TV; 1.		
DR	PROSITE; PS00484; THYROGLOBULIN_1; 1.		
KW	Glycoprotein; MHC; Proteoglycan; Transmembrane; Alternative splicing;		
KW	Signal-anchor.		
FT	DOMAIN 1 30	CYTOSOLASMIC (POTENTIAL).	
FT	TRANSMEM 31 56	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)	

FT DOMAIN 57 280 EXTRACELLULAR (POTENTIAL).  
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 266 266 O-LINKED (GLYCOSAMINOGLYCAN) (BY  
 SIMILARITY).  
 FT DOMAIN 212 255 THYROGLOBULIN TYPE I.  
 FT VARSPIC 193 256 MISSING (IN SHORT ISOCORM).  
 SO SEQUENCE 280 AA: 31642 MW: D935D169A98B5732 CRC64:

Query Match 11.9%, Score 153, DB 1: Length 280;  
 Best Local Similarity 38.4%, Pred. No. 1e-06;  
 Matches 33; Conservative 10; Mismatches 35; Indels 8; Gaps 1;

OY 19 WVFIFCAITTEAS-----LTKCOQLASANSGLIGYVPOCKEKEGEPEKQCMGS 70  
 DB 171 WLFEMSKNSLEKQPTQTPRVVLKCOEYSHLPDVHPRGAFRKCDEKNQNYPLQCHS 230

OY 71 TGYCVCYDEGCKEILGTIKRSPDCS 96  
 DB 231 TGYCVCYFPNGTEVPHTSRGRHNS 256

## RESULT 12

ID NIDO\_HUMAN STANDARD: PRT: 1247 AA.

AC p14563: Q1942; (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NIDODGEN PRECURSOR (ENTACTIN).  
 GN NID.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID:9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:90091745; PubMed:2574658;  
 RA Nagayoshi T., Sanborn D., Hickok N.J., Olsen D.R., Fazio M.J.,  
 Chu M.-L., Knowlton R., Mann K., Deutzmann R., Timpl R., Uitto J.;  
 RT "Human nidogen: complete amino acid sequence and structural domains  
 deduced from cDNAs," and evidence for polymorphism of the gene,"  
 RL DNA 8:581-594(1989).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:96044428; PubMed:7557988;  
 RA Zimmermann K., Holschen S., Hainer M., Nischt R.;  
 RT "Genomic sequences and structural organization of the human nidogen  
 gene (NID).";  
 RL Genomics 27:245-250(1995).

RN [3]  
 RP SEQUENCE OF 667-1247 FROM N.A.

RC TISSUE=Placenta;  
 RX MEDLINE:89270475; PubMed:2471408;  
 RA Olsen D.R., Nagayoshi T., Fazio M., Mattei M.-G., Passage E., Weil D.,  
 Timpl R., Chu M.-L., Uitto J.;  
 RT "Human nidogen: cDNA cloning, cellular expression, and mapping of the  
 RT gene to chromosome 19q43.";  
 RL Am. J. Hum. Genet. 44:876-885(1989).  
 CC -1- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN  
 CC BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.  
 CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-  
 CC EXTRACELLULAR MATRIX INTERACTIONS.  
 CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.  
 CC -1- PTM: N- AND O-GLYCOSYLATED.  
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M30269; AAA59932.1; -;  
 CC EMBL: X82245; CA557709.1; -;  
 CC EMBL: X84819; CA557709.1; JOINED.  
 CC EMBL: X84820; CA557709.1; JOINED.  
 CC EMBL: X84821; CA557709.1; JOINED.  
 CC EMBL: X84822; CA557709.1; JOINED.  
 CC EMBL: X84823; CA557709.1; JOINED.  
 CC EMBL: X84824; CA557709.1; JOINED.  
 CC EMBL: X84825; CA557709.1; JOINED.  
 CC EMBL: X84826; CA557709.1; JOINED.  
 CC EMBL: X84827; CA557709.1; JOINED.  
 CC EMBL: X84828; CA557709.1; JOINED.  
 CC EMBL: X84829; CA557709.1; JOINED.  
 CC EMBL: X84830; CA557709.1; JOINED.  
 CC EMBL: X84831; CA557709.1; JOINED.  
 CC EMBL: X84832; CA557709.1; JOINED.  
 CC EMBL: X84833; CA557709.1; JOINED.  
 CC EMBL: X84834; CA557709.1; JOINED.  
 CC EMBL: X84835; CA557709.1; JOINED.  
 CC EMBL: X84836; CA557709.1; JOINED.  
 CC EMBL: X84837; CA557709.1; JOINED.  
 CC EMBL: M27445; AAA57261.1; -;  
 CC PIR: A33322; MMRND.  
 CC HSSP: P07204; 2ADX.  
 CC MIM: 131390; -;  
 CC InterPro: IPR000152; Asx\_hydroxyl.  
 CC InterPro: IPR000561; EGF-like.  
 CC InterPro: IPR001881; EGF Ca.  
 CC InterPro: IPR000033; Ldl\_rcptor\_rep.  
 CC InterPro: IPR003886; Nidogen\_ext.  
 CC InterPro: IPR00716; Thyroglobulin\_1.  
 CC Pfam: PF00008; EGF; 6.  
 CC Pfam: PF00058; Ldl\_recept\_b; 3.  
 CC Pfam: PF00086; Thyroglobulin\_1; 1.  
 CC SMART: SM00179; EGF\_CA; 2.  
 CC SMART: SM00001; EGF\_Like; 4.  
 CC SMART: SM00135; LY; 5.  
 CC SMART: SM00539; NIDO; 1.  
 CC SMART: SM00211; TY; 1.  
 CC PROSITE: PS00010; ASX\_HYDROXYL; 3.  
 CC PROSITE: PS00484; THYROGLOBULIN\_1; 1.  
 CC PROSITE: PS00022; EGF\_1; 1.  
 CC PROSITE: PS01186; EGF\_2; 5.  
 CC PROSITE: PS01187; EGF\_CA; 2.  
 CC KW Basement membrane; Extracellular matrix; Glycoprotein; Sulfation;  
 CC KW Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion.  
 CC FT SIGNAL 1 28  
 CC FT CHAIN 29 1247 NIDODGEN.  
 CC FT DOMAIN 29 669 I (LARGER GLOBULAR DOMAIN).  
 CC FT DOMAIN 670 917 II (CYSTEINE-RICH).  
 CC FT DOMAIN 918 1247 III (SMALLER GLOBULAR DOMAIN).  
 CC FT DOMAIN 386 426 EGF-LIKE 1.  
 CC FT DOMAIN 668 709 EGF-LIKE 2.  
 CC FT DOMAIN 710 751 EGF-LIKE 3.  
 CC FT DOMAIN 752 801 EGF-LIKE 4.  
 CC FT DOMAIN 802 840 EGF-LIKE 5.  
 CC FT DOMAIN 872 919 THYROGLOBULIN TYPE I.  
 CC FT DOMAIN 989 1030 LDL-RECEPTOR YWTD MOTIF 1.  
 CC FT DOMAIN 1032 1073 LDL-RECEPTOR YWTD MOTIF 2.  
 CC FT DOMAIN 1075 1118 LDL-RECEPTOR YWTD MOTIF 3.  
 CC FT DOMAIN 1124 1163 LDL-RECEPTOR YWTD MOTIF 4.  
 CC FT DOMAIN 1208 1244 EGF-LIKE 6.  
 CC FT MOD\_RES 289 289 SULFATION (POTENTIAL).  
 CC FT MOD\_RES 296 296 SULFATION (POTENTIAL).  
 CC FT DISULFID 672 685 BY SIMILARITY.  
 CC FT DISULFID 679 695 BY SIMILARITY.  
 CC FT DISULFID 708 708 BY SIMILARITY.  
 CC FT DISULFID 714 727 BY SIMILARITY.

FT DISULFID 721 736 BY SIMILARITY.  
 FT DISULFID 736 750 BY SIMILARITY.  
 FT DISULFID 762 777 BY SIMILARITY.  
 FT DISULFID 769 787 BY SIMILARITY.  
 FT DISULFID 789 800 BY SIMILARITY.  
 FT DISULFID 806 817 BY SIMILARITY.  
 FT DISULFID 811 826 BY SIMILARITY.  
 FT DISULFID 828 839 BY SIMILARITY.  
 FT DISULFID 1212 1223 BY SIMILARITY.  
 FT DISULFID 1219 1232 BY SIMILARITY.  
 FT DISULFID 1234 1243 BY SIMILARITY.  
 FT SITE 702 704 CELL ATTACHMENT SITE.  
 FT CARBOHYD 1137 1137 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 33 34 EL -> SS (IN REF. 2).  
 FT CONFLICT 37 42 RCPGOG -> SAPDR (IN REF. 2).  
 FT CONFLICT 1115 1115 H -> T (IN REF. 3).  
 SQ SEQUENCE 1247 AA; 136488 MW; 4681B5B3CEC1575B CRC64;

Query Match 11.9%; Score 153; DB 1; Length 1247;  
 Best Local Similarity 25.5%; Pred. No. 4.6e-06;  
 Matches 56; Conservative 18; Mismatches 70; Indels 76; Gaps 11;

QY 25 CAITSTASLTKCQOLASNSGLITVPP--QCKETGEFEKQCGSTGYCWCV--D 78  
 DB 750 CVAVVDPRIINCE-----TGLHNCIDIPORAQCIYTG-----GSSYTCSCLPFS 794  
 QY 79 EDGKEILGTRKRGSPDCSRKALITLCOMQALITVPPMCGPSCKADGSDVEQCCAS 138  
 DB 795 GDGQ-----ACQDVE-----CQPSRCHPDAR-----CYNT 820  
 QY 139 NGECYCVDRKKELEGTROOGRPTC--ERHISECEARIKAH-----SNSLRVE 185  
 DB 821 PGSTCCCKRGYGDGR-----CVPEVEKTRCOHER--EHLIGAAGATDPQRIIPG 872  
 QY 186 MFVEPCLEDSYNVQCWPTGYCVCVDEGCVAVPQSDVR 225  
 DB 873 LFVEPCDAHGYAPTOCHGSTGYCVCVDRGREVETRR 912

RESULT 13  
 HG2A\_HUMAN STANDARD; PRT; 296 AA.  
 AC P04233; Q29832;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, GAMMA CHAIN (HLA-DR ANTIGENS  
 DE ASSOCIATED INVARIANT CHAIN) (P33) (CD74 ANTIGEN).  
 GN CD74 OR DILAG.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=86093681; PubMed=3001652;  
 RA Kudo J., Chao L.-Y., Narni F., Saunders G.F.;  
 RA "Structure of the human gene encoding the invariant gamma-chain of  
 RA class II histocompatibility antigens.";  
 RL Nucleic Acids Res. 13:8837-8841(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=84207945; PubMed=6586420;  
 RA Strubin M., Mach B., Long E.O.;  
 RA "The complete sequence of the mRNA for the HLA-DR-associated  
 RA invariant chain reveals a polypeptide with an unusual transmembrane  
 RA polarity.";  
 RL EMBO J. 3:869-872(1984).  
 RN [3]  
 RP SEQUENCE OF 27-96 FROM N.A.  
 RX MEDLINE=84170234; PubMed=6324166;  
 RX Claesson L., Larhammar D., Rask L., Peterson P.A.;

RT \*cDNA clone for the human invariant gamma chain of class II  
 RT histocompatibility antigens and its implications for the protein  
 RT structure.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:7395-7399(1983).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86233451; PubMed=3459184;  
 RA O'Sullivan D.M., Larhammar D., Wilson M.C., Peterson P.A.,  
 RA Quaranta V.;  
 RA "Structure of the human Ia-associated invariant (gamma)-chain gene:  
 RA identification of 5' sequences shared with major histocompatibility  
 RA complex class II genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4484-4488(1986).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.75 ANGSTROMS) OF 103-117.  
 RX MEDLINE=96085023; PubMed=7477400;  
 RA Ghosh P., Amaya M., Mellins E., Wiley D.C.;  
 RA "The structure of an intermediate in class II MHC maturation: CLIP  
 RT bound to HLA-DR3.";  
 RL Nature 378:457-462(1995).  
 RN [6]  
 RP STRUCTURE BY NMR OF 134-208.  
 RX MEDLINE=95059718; PubMed=843486;  
 RA Jasnoff A., Wagner G., Wiley D.C.;  
 RA "Structure of a trimeric domain of the MHC class II-associated  
 RT chaperonin and targeting protein I1.";  
 RL EMBO J. 17:6812-6818(1998).  
 CC - FUNCTION: PLAYS A CRITICAL ROLE IN MHC CLASS II ANTIGEN PROCESSING  
 CC BY STABILIZING PEPTIDE-FREE CLASS II ALPHA/BETA HETERODIMERS IN A  
 CC COMPLEX SOON AFTER THEIR SYNTHESIS AND DIRECTING TRANSPORT OF THE  
 CC COMPLEX FROM THE ENDOPLASMIC RETICULUM TO COMPARTMENTS WHERE  
 CC PEPTIDE LOADING OF CLASS II TAKES PLACE.  
 CC - SUBUNIT: NONAMER COMPOSED OF THREE ALPHA/BETA/GAMMA HETEROTIMERS.  
 CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).  
 CC - SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.  
 CC - DATABASE: NAME=PROIM; NOTE=CD guide CD74 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/ptow/cd/cd74.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X03339; CAA27046.1; -;  
 CC EMBL; X03340; CAA27047.1; -;  
 CC EMBL; K01144; AAA36304.1; -;  
 CC EMBL; X00497; CAA25192.1; -;  
 CC EMBL; X00497; CAA25193.1; -;  
 CC F1R; A30060; HUHG.  
 CC PDB; 1A6A; 17-MAY-98.  
 CC PDB; 11E; 15-FEB-99.  
 CC MIM; 142790; -;  
 CC InterPro: IPR000716; Thyroglobulin\_1.  
 CC Pfam: PF00086; thyroglobulin\_1.  
 DR SMART; SM00211; TV, 1.  
 DR PROSITE; PS00484; THYROGLOBULIN\_1; 1.  
 KW Glycoprotein; MHC; Proteoglycan; Transmembrane; Alternative splicing;  
 KW Signal-anchor; 3D-structure.  
 FT DOMAIN 1 46 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 47 72 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT 73 296 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 103 117 CLIP.  
 FT CARBOHYD 130 136 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 282 282 O-LINKED (GLYCOSAMINOGLYCAN).  
 FT DOMAIN 228 271 THYROGLOBULIN TYPE I.  
 FT VARSPPLIC 209 272 MISSING (IN SHORT ISOFORM).  
 FT CONFLICT 167 167 T -> R (IN REF. 2 AND 3).  
 SQ SEQUENCE 296 AA; 33460 MW; C9A301208D5FB918 CRC64;

[illegible]

```

OY      141 ECVCKR-----KGELEGRFOOGRP-----TCRRHSECEERIRAHNSNR 183
Db      137 K-CLOKHLAKINDRSTSGCKMKYIGAPREARIPVPOSSQSELHRALE-RLAASQSPRH 193

OY      184 VEFN---VEPCLEDGSGYNPVOCMP-----TGYCWD-EGSVKPKG 221
Db      194 EDLYIPPCDRNGNFHFKCHPLDGOGRGKCMVCVKRTGKVLPG 239

RESULT 15
IBP4_MOUSE
ID      IBP4_MOUSE          STANDARD:          PRT:          254 AA.
AC      P47879: 036666;
DT      01-FEB-1998 (Rel. 33, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 4, PRECURSOR (IGFBP-4)
DE      (IBP-4) OR IGFBP-4.
GN      IGFBP4 OR IGFBP-4.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxId=10090;
[1]
RN      RP      SEQUENCE FROM N.A.
RC      TISSUE=Liver;
RX      MEDLINE=95121750; PubMed=7529732;
RA      Schuller A.G.P., Grotten C., van Neck J.W., Zwarthoff E.C.,
RA      Drop S.D.S.;
RT      "cDNA cloning and mRNA expression of the six mouse insulin-like
RT      growth factor binding proteins.";
RL      Mol. Cell. Endocrinol. 104:57-66(1994).
[2]
RN      RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6J;
RX      MEDLINE=95152444; PubMed=7531538;
RA      Bethel C.R., Vitullo J.C., Miller R.E., Aron D.C.;
RT      "Molecular cloning of mouse insulin-like growth factor binding
RT      protein 4 (IGFBP4) cDNA and expression of a fusion protein with
RT      IGF-binding activity.";
RL      Biochem. Mol. Biol. Int. 34:385-392(1994).
[3]
RN      RP      SEQUENCE OF 1-114 FROM N.A.
RC      STRAIN=BALB/C; TISSUE=Brain;
RX      MEDLINE=98127740; PubMed=9468222;
RA      Gianschmid H., Varga F., Luemgway E., Klaushofer K.;
RT      "Characterization of the mouse insulin-like growth factor binding
RT      protein 4 gene regulatory region and expression studies.";
RL      DNA Cell Biol. 17:51-60(1998).
-1-    FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs
CC      AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC      PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE
CC      INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.
-1-    SUBCELLULAR LOCATION: SECRETED.
CC      -1- MISCELLANEOUS: BINDS IGF-II MORE THAN IGF-I.
CC      -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC      -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC      PROTEIN FAMILY.

```

```
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb.sib.ch).
CC      -----
DR      EMBL; X81582; CAA57272.1; -
DR      EMBL; X76066; CAA53667.1; -
DR      EMBL; Z95492; CAB08859.1; -
DR      HSP; P01180; INPO.
DR      MGD; MGI:96439; Igfbp4.
DR      InterPro; IPR000867; IGFBP.
DR      InterPro; IPR000716; Thyroglobulin_1.
DR      Pfam; PF00219; IGFBP; 1.
DR      Pfam; PF00086; thyroglobulin_1; 1.
DR      SMART; SM00121; IB; 1.
DR      SMART; SM00211; TY; 1.
DR      PROSITE; PS00222; IGF_BINDING; 1.
DR      PROSITE; PS00484; THYROGLOBULIN_1; 1.
DR      Growth factor binding; signal; Glycoprotein.
KW      Growth factor binding; signal; Glycoprotein.
FT      SIGNAL      1      21      BY SIMILARITY.
FT      CHAIN       22      254      INSULIN-LIKE GROWTH FACTOR BINDING
FT                                     PROTEIN 4.
FT      DOMAIN      196      245      THYROGLOBULIN TYPE I.
FT      CARBOHYD     125      125      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CONFLICT     4         5      FG -> CS (IN REF. 1).
FT      CONFLICT    13        13      A -> T (IN REF. 1).
FT      CONFLICT    56        56      C -> S (IN REF. 1).
FT      CONFLICT    67        67      C -> G (IN REF. 1).
SQ      SEQUENCE    254 AA; 27807 MW; 58EF89CB514AE17 CRC64;

Query Match      10.1%; Score 129.5; DB 1; Length 254;
Best Local Similarity 24.2%; Pred. No. 0.00012;
Matches 52; Conservative 34; Mismatches 76; Indels 53; Gaps 12;

OY      37 CQQLQASANSGL-----IGTVYPCKETGEFEKQCKWSTGYCWCVEDGKEIL 85
DB      44 CEELVREPCCGCATCATLGLGMPGCGVTPRCG-----SGMRCYPPRGV-----EKPL 90
OY      86 GFKINGSPDCSRKALFLCQMMQALIVNPGWCGPSCKADGSPDEVQCCASNGECY-- 143
DB      91 RFLMHGQGYC-----TELSEIFAIDSLQTSKDSESEHPNNSFN--PCSAHDHRCLOK 141
OY      144 ----CVDKKGELEGT-RQGRP-----TCERHLSCEEARIKAHNSLRLVEMF--VPEC 191
DB      142 HMAKIRDRKMKIIVGTPREPRPVQSGQSELHRALE-RLAASQSRTHEDLFILIPINC 200
OY      192 LEDGSYNPQWPS----TGYCWVD-EGVKVPG 221
DB      201 DRNGNFHPKQCHPALDGGKCKMCVDRKTVKLP 235
```

Search completed: March 8, 2002, 09:25:29  
Job time: 192 sec

4  
.  
.  
.  
.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2002, 09:24:02 ; Search time 23.53 Seconds

(Without alignments)  
1435.993 Million cell updates/sec

Title: US-09-445-480A-2

Sequence: 1 MALSQNAQAFKSGFVMIWV.....VDEGGVKGPSDVRKRPPTC 231

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: SP\_ARCHA:\*

2: SP\_BACTERIA:\*

3: SP\_FUNGI:\*

4: SP\_HUMAN:\*

5: SP\_INVERTEBRATE:\*

6: SP\_MAMMAL:\*

7: SP\_MHC:\*

8: SP\_ORGANELLE:\*

9: SP\_PHAGE:\*

10: SP\_PLANT:\*

11: SP\_PROTIST:\*

12: SP\_VIRUS:\*

13: SP\_VIRUS:\*

14: SP\_UNCLASSIFIED:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1287	100.0	231	5 Q9U6K8	Q9U6K8 actinia equ
2	1278	99.3	231	5 Q9U4R8	Q9U4R8 actinia equ
3	1279	21.7	212	6 Q28330	Q28330 capra hircu
4	259.5	20.2	446	11 Q9EP95	Q9EP95 mus musculu
5	259	20.1	457	4 Q9H300	Q9H300 homo sapien
6	255.5	19.6	452	11 Q9H4F7	Q9H4F7 mus musculu
7	252.5	19.6	452	11 Q9H4F8	Q9H4F8 mus musculu
8	252.5	19.6	452	4 Q9H4F7	Q9H4F7 mus musculu
9	236.5	18.5	1161	5 Q9H901	Q9H901 naucyrtina
10	177.5	13.8	88	5 Q9H900	Q9H900 salvelinus
11	177.5	13.4	613	5 Q9H901	Q9H901 salvelinus
12	171.5	13.3	1511	5 Q9H901	Q9H901 salvelinus
13	167.5	13.0	126	4 Q9H901	Q9H901 salvelinus
14	167	13.0	234	7 Q9H901	Q9H901 salvelinus
15	167	13.0	237	13 Q9H901	Q9H901 salvelinus
16	166.5	12.9	279	4 Q9H4F7	Q9H4F7 mus musculu
17	165	12.8	2225	5 Q45881	Q45881 caenorhabdi
18	161	12.5	439	4 Q9H901	Q9H901 salvelinus
19	160.5	12.5	474	13 Q9H901	Q9H901 salvelinus

20	159.5	12.4	286	7 Q9BD54	Q9BD54 gallus gall
21	159	12.4	442	11 Q62288	Q62288 mus musculu
22	158.5	12.3	2150	5 Q44131	Q44131 caenorhabdi
23	156	12.1	433	4 Q9B016	Q9B016 homo sapien
24	154	12.0	436	11 Q9ER59	Q9ER59 mus musculu
25	153.5	11.9	198	13 Q9PVD9	Q9PVD9 brachydantio
26	153	11.9	436	4 Q75705	Q75705 homo sapien
27	150	11.7	71	7 Q78208	Q78208 homo sapien
28	145	11.3	424	4 Q92563	Q92563 mus musculu
29	143	11.1	423	11 Q9ER58	Q9ER58 mus musculu
30	143	11.1	660	5 Q9VCR3	Q9VCR3 drosophila
31	128.5	10.0	505	5 Q25431	Q25431 lytechinus
32	118.5	9.2	227	13 Q9D612	Q9D612 coturnix co
33	114.5	8.9	1187	2 Q49549	Q49549 mycoplasma
34	114	8.9	911	11 Q9CRX6	Q9CRX6 mus musculu
35	112.5	8.7	226	5 Q44466	Q44466 caenorhabdi
36	112	8.7	369	5 Q17457	Q17457 caenorhabdi
37	111.5	8.7	949	5 P90956	P90956 caenorhabdi
38	110.5	8.6	291	6 Q9G1V5	Q9G1V5 bos taurus
39	110.5	8.6	434	5 Q9W0A0	Q9W0A0 drosophila
40	109.5	8.5	293	6 Q9PT10	Q9PT10 sus scrofa
41	109.5	8.5	1069	5 Q9BPS2	Q9BPS2 bombyx mori
42	108	8.4	955	4 Q9H920	Q9H920 drosophila
43	108	8.4	1284	5 Q9H920	Q9H920 drosophila
44	108	8.4	1664	5 Q9H920	Q9H920 drosophila
45	108	8.4	1999	4 Q9H940	Q9H940 homo sapien

## ALIGNMENTS

RESULT ID	1	PREDIMINARY:	PRF:	231 AA.
Q9U6K8	Q9U6K8			
AC	Q9U6K8			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	ECUISTATIN PRECURSOR.			
GN	EQ1.			
OS	Actinia equina (European sea anemone).			
OC	Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;			
CC	Nynanthozoa; Actiniidae; Actinia.			
OX	NCBI_TaxID-6106;			
RN	11			
RP	SEQUENCE FROM N.A.			
RA	Galea K., Strudel B., Bayec S., Turk V., Lenarcic B.;			
RT	"Molecular cloning and functional expression of ecuistatin.";			
RL	Biol. Chem. 0:0-0(1999).			
DR	EMBL: AF184891; AAF02722.1; "			
DR	InterPro: IPR000716; Thyroglobulin_1.			
DR	Pfam: PF00086; thyroglobulin_1; 3.			
DR	PROSITE: PS00484; THYROGLOBULIN_1; 2.			
DR	SMART: SM00211; Ty; 3.			
KW	SM00211.			
FT	SIGNAL.			
FT	CHAIN			
FT	33			
FT	231			
FT	POTENTIAL.			
FT	ECUISTATIN.			
FT	SEQUENCE			
FT	231 AA; 25406 MW; E625D49272835958 CMC64;			
Query Match	100.0%; Score 1287; DB 5; Length 231;			
Best Local Similarity	100.0%; Pred. No. 6.1e-129; Indels 0; Gaps 0;			
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 MALSQNAQAFKSGFVMIWVLFACATITSTASITKCOLOLSANSGLITGYPOCKEKG 60			
DB	1 MALSQNAQAFKSGFVMIWVLFACATITSTASITKCOLOLSANSGLITGYPOCKEKG 60			
QY	61 EPEKCKWGSIGYCWYDEDEKRLITKIRSPKSRRAALITLCOMKQAIYVWPGMG 120			
DB	61 EPEKCKWGSIGYCWYDEDEKRLITKIRSPKSRRAALITLCOMKQAIYVWPGMG 120			
QY	121 PPSCKADSGFDEVCAGSNGECYVDKRGKELEGTROGGRPTCERHUSSECEARIRAHNS 180			

```

DB 121 PSCADSGFDEVOCCASNGECYCVDKKKELEGTRKROGRTCEHLSSECEBARIKAHSN 180
QY 181 SLRVEFVPECLDEGSGYNVOCWPCSTGYCWCDEGAVVPESDVRFRKPTC 231
DB 181 SLRVEFVPECLDEGSGYNVOCWPCSTGYCWCDEGAVVPESDVRFRKPTC 231

RESULT 2
ID Q904R8 PRELIMINARY: PRT: 231 AA.
AC Q904R8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE EQUITATIN PRECURSOR.
OS Actinia equina (European sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;
OC Nyantheae; Actiniidae; Actinia.
OX NCBI_Taxid=6106;
RN [1]
RP SEQUENCE FROM N.A.
RA Strukej B., Pungercar J., Jongsma M.A.;
RT "Primary structure of equistatin deduced from cDNA.";
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF156179; AAF24173.1;
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00086; thyroglobulin_1.
DR PROSITE: PS00484; THYROGLOBULIN_1; 2.
DR SMART: SM00211; TY; 3.
KW SIGNAL.
FT SIGNAL.
SQ SEQUENCE 231 AA; 25410 MW; 07E256877D3356E9 CRC64;

Query Match
Best Local Similarity 99.3%; Score 1278; DB 5; Length 231;
Matches 229; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MALSONAKRSKGFVMTVFTICATSTPEASTKRCQLOASNSGLIGTVPOCKETG 60
DB 1 MALSONAKRSKGFVMTVFTICATSTPEASTKRCQLOASNSGLIGTVPOCKETG 60
QY 61 EPEKQCGSTGYCWCVDGDEGKELTGKINGSDCSRRKALITLCOMQAIIVNPGMG 120
DB 61 EPEKQCGSTGYCWCVDGDEGKELTGKINGSDCSRRKALITLCOMQAIIVNPGMG 120
QY 121 PSCADSGFDEVOCCASNGECYCVDKKKELEGTRKROGRTCEHLSSECEBARIKAHSN 180
DB 121 PSCADSGFDEVOCCASNGECYCVDKKKELEGTRKROGRTCEHLSSECEBARIKAHSN 180
QY 181 SLRVEFVPECLDEGSGYNVOCWPCSTGYCWCDEGAVVPESDVRFRKPTC 231
DB 181 SLRVEFVPECLDEGSGYNVOCWPCSTGYCWCDEGAVVPESDVRFRKPTC 231

RESULT 3
ID Q28330 PRELIMINARY: PRT: 212 AA.
AC Q28330;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE DNA FOR THYROGLOBULIN (TG) (FRAGMENT).
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_Taxid=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=69207610; PubMed=2495821;
RA Van Ommen G.J.B., de Vijlder J.J.M., Sterk A., Mercken L.O.Y.,

```

```

RA Arbery A.C., Baas F.;
RT "Studies on the structures of the normal and abnormal goat
RT thyroglobulin genes.";
RL Biochimie 71:211-221(1989).
DR EMBL: X14997; CA433105.1;
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00086; thyroglobulin_1; 2.
DR PROSITE: PS00484; THYROGLOBULIN_1; 2.
DR SMART: SM00211; TY; 2.
DR NON TER 212
FT NON TER 212
SQ SEQUENCE 212 AA; 23791 MW; 100025E08E9B9E8 CRC64;

Query Match
Best Local Similarity 21.7%; Score 279; DB 6; Length 212;
Matches 62; Conservative 18; Mismatches 60; Indels 18; Gaps 6;

QY 52 YVPOCKETGEFEKQCGSTGYCWCVDGDEGKELTGKINGSDCSRRKALITLCOMQAI 111
DB 48 YVPOCAEBSFQYVQCGCKDASGCVADAGREVPGSRQGRP-----ACLSFCQLQKQ 102
QY 112 I-----VNVPGMGCPSCADSGFDEVOCCASNGECYCVDKKKELEGTRKROGRT-CEH 165
DB 103 ILSSYINSTASTVLPQCQDSGDYSFVQCGLRRKQCGVADAGMEVYCTQGRPRQPCR 162
QY 166 HLESECE-EARIKAHNSLRVEFVPECLDEGSGYNVOC 202
DB 163 ----SCIRNRRLHGVGDRSP---PQSPDAFRPQC 194

RESULT 4
ID Q9895 PRELIMINARY: PRT: 446 AA.
AC Q9895;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SECRETED MODULAR CALCIUM-BINDING PROTEIN 2 PRECURSOR.
GN SMOG2 OR SMOG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Vanname C., Goessling S., Maurer P.;
RT "Cloning of smoc-2, a secreted modular calcium-binding protein.";
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
DE -1- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL: AJ249901; CAC10355.1;
DR MGI: MGI:1929881; Smoc2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR002350; kazal.
DR Pfam: PF00036; ehand; 2.
DR Pfam: PF00050; kazal; 1.
DR Pfam: PF00086; thyroglobulin_1; 2.
DR SMART: SM00280; KAZAL; 1.
DR SMART: SM00211; TY; 2.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR Calcium-binding; Signal.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 21
SQ SEQUENCE 446 AA; 49818 MW; 6C3303A67212CC2A CRC64;

Query Match
Best Local Similarity 20.2%; Score 259.5; DB 11; Length 446;
Matches 58; Conservative 30; Mismatches 86; Indels 27; Gaps 5;

QY 41 QASANGSLGTVYVQCKETGEFEKQCGSTGYCWCVDGDEGKELTGKTRKSPDCSRRK 99

```

```
Db 98 QEQARKFEQVFEPCNDGTYGVOCHSTGYCMCVTPNGRPISTGTAVAHKTPRCP--- 154
Qy 100 AALTLQMMQAIIVNPGMCGS-----PSCKADGSFDEYQCCASNGECYCDKKGKE 151
Db 155 -----GSINKVYQREBAGKADAAPALETPQGDDEDIASRYPLILMTQYVKSQ 205
Qy 152 LEGRQGRPTCERHLSCEEARIKAHNSILRVEMFPECELEDGSYNPVOCWPGSTGYCWC 211
Db 206 NKTNNKNSASCDQEHQSALEBAKQPKNDN-----VVIPECAHGGLYKPVQCHPSTGYCWC 260
Qy 212 V-DEGVKVPGSVDVRFKRPCT 231
Db 261 VLVDTPGRIPIGTSTRYEQPKC 281

RESULT 5
Q9H3U0 PRELIMINARY; PRT; 457 AA.
AC Q9H3U0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SNAP-2B.
GN SNAP-2B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA Nishimoto S., Toyoda H., Tawara J., Aoki T., Komurasaki T.;
RT "Molecular cloning and characterization of human smooth muscle cell
RT associated protein-2b(SNAP-2b).";
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL: AB014737; BAB20274.1; -.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR002350; Kazal.
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00036; ehand; 2.
DR Pfam: PF00050; kazal; 1.
DR Pfam: PF00086; thyroglobulin_1; 2.
DR SMART: SM00280; KAZAL; 1.
DR SMART: SM00211; TY; 2.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
KW Calcium-binding.
SQ SEQUENCE 457 AA; 50862 MW; D20AE3BB82785925 CRC64;
```

```
Query Match 20.1%; Score 259; DB 4; Length 457;
Best Local Similarity 29.6%; Pred. No. 2.7e-19;
Matches 60; Conservative 32; Mismatches 91; Indels 20; Gaps 7;
```

```
ID Q9WVN9 PRELIMINARY; PRT; 452 AA.
AC Q9WVN9;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SPARC-RELATED PROTEIN.
GN SMOC1 OR SRG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RA Poliev A., Plachov D.;
RT "A Sparc-related gene (SRG).";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF070470; MAD41590.1; -.
DR HSSP: P09486; ISRA.
DR MGI:1929878; Smoc1.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR002350; Kazal.
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00050; kazal; 1.
DR Pfam: PF00086; thyroglobulin_1; 2.
DR SMART: SM00280; KAZAL; 1.
DR SMART: SM00211; TY; 2.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
DR PROSITE: PS00484; THYROGLOBULIN_1; 2.
SQ SEQUENCE 452 AA; 49763 MW; EAF73CB0AFC060F0 CRC64;
```

```
Query Match 19.9%; Score 255.5; DB 11; Length 452;
Best Local Similarity 32.0%; Pred. No. 6.3e-19;
Matches 71; Conservative 25; Mismatches 87; Indels 39; Gaps 8;
```

```
Qy 31 EASLTKGCO---QLQASNSGLIGTVPOCKEFGEEKQCGSGYCMCVDEDEKELIG 86
Db 88 DAGQSKRLERAOALEQAKKPOEAVVPPCGEDGSFTYQCHTYGTCMCTPPDGKPISG 147
Qy 87 TKIRG-SPPCSRRKAALTLQMMQAIIVNPGMCGPPSCADGS-----FDEVO 134
Db 148 SSVQNTKTPVCSGQVTDKPLSQ-----GNSGRKDDGSKPTPTMETQPVFDDGE 194
Qy 135 CCASNECYCVDKKKELEGTROQGRP---TGER-HLSCEARIKAHNSILRVEMFPE 190
Db 195 ITAPTLMIKHLVTKDKSLNNTVNRSEKVSQDOEROSALEBAR-----QNPREGIVPE 249
Qy 191 CLEDGSYNPVOCWPGSTGYCMCV-DEGVKVPGSVDVRFKRPCT 231
Db 250 CARGGLKPVQCHQSTGYCMCVLVDTPGRIPIGTSTRYVWPSC 291
```

```
RESULT 7
Q9H4F8 PRELIMINARY; PRT; 434 AA.
AC Q9H4F8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SECRETED MODULAR CALCIUM-BINDING PROTEIN PRECURSOR.
GN SMOC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Vannanme C., Goesling S., Maurer P.;
RT "Cloning of human smoc-1, a novel secreted modular calcium-binding
RT protein.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
```

DR EMBL: AJ249900; CAC10352.1; -.  
DR InterPro: IPR002048; EF-hand.  
DR InterPro: IPR002350; kazal.  
DR InterPro: IPR00716; thyroglobulin\_1.  
DR Pfam: PF00050; kazal; 1.  
DR Pfam: PF00086; thyroglobulin\_1; 2.  
DR SMART: SM00280; KAZAL; 1.  
DR SMART: SM00211; TY; 2.  
DR PROSITE: PS00018; EF\_HAND; UNKNOWN\_1.  
KW SIGNAL.  
FT CHAIN 1 26 POTENTIAL.  
FT CHAIN 27 434 SMO-1.  
SQ SEQUENCE 434 AA; 48162 MW; 2CB639212BA42478 CRC64;

Query Match 19.6%; Score 252.5; DB 4; Length 434;  
Best Local Similarity 31.7%; Pred. No. 1.3e-18;  
Matches 70; Conservative 26; Mismatches 88; Indels 37; Gaps 8;

OY 31 EASLTKCQ---QLQASNSGLITGYVPCKETGEFEKQCGWSTGYCWCVDGDKETLG 86  
DB 89 DAGSKCHLERQALFQAKKQEAIVFECGEDGSFTYVOCHTITGYCVCVTPDGKPISG 148  
OY 87 TKIRG-SPDCSRKALTLQOMQAIIVNPGWCGPPSCADGS-----FDEVQ 134  
DB 149 SSVQKKTTPVCSGVTDKPLSQ-----GNSGRKDDGSKPPTMETQPVFDGDE 195  
OY 135 CCASNGECYCDKKKELEGTROGGRP---TCERHLSCECEARIKAHNSLRVEMFVPEC 191  
DB 196 ITAPLMTIKHLVTKDSKLNNTNINSEKVSQDQ---ERQSALEDAQNP-REGIVIPFC 251  
OY 192 LEDGSYNPVOCWPTGYCWCY-DEGGVAVPGSDVFRKRPYC 231  
DB 252 AFGILKRYQCHQSTGYCWCYLVDTGRPLPSTIRYVNPSC 292

RESULT 8  
O9H3U7 PRELIMINARY: PRT: 446 AA.  
AC O9H3U7.  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE SMAP-2.  
GN SMAP-2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HEART;  
RA Nishimoto S., Toyoda H., Tawara J., Aoki T., Komurasaki T.;  
RT "Molecular cloning and characterization of human smooth muscle cell  
associated protein-2 (SMAP-2).";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO EF-HAND FAMILY.  
CC EMBL: AB014730; BAB20267.1; -.  
DR InterPro: IPR002048; EF-hand.  
DR InterPro: IPR002350; kazal.  
DR InterPro: IPR00716; thyroglobulin\_1.  
DR Pfam: PF00036; ehand; 2.  
DR Pfam: PF00050; kazal; 1.  
DR Pfam: PF00086; thyroglobulin\_1; 2.  
DR SMART: SM00280; KAZAL; 1.  
DR SMART: SM00211; TY; 2.  
DR PROSITE: PS00018; EF\_HAND; UNKNOWN\_1.  
KW Calcium-binding.  
SQ SEQUENCE 446 AA; 49684 MW; 89D6FEF1F8F0575B CRC64;

Query Match 19.6%; Score 252.5; DB 4; Length 446;  
Best Local Similarity 26.8%; Pred. No. 1.3e-18;

Matches 56; Conservative 29; Mismatches 81; Indels 43; Gaps 5;  
OY 41 QASNSGLITGYVPCKETGEFEKQCGWSTGYCWCVDGDKETLTKI-RSPDCSRK 99  
DB 98 QEQARKEEQGVFIPECDDSTYSVOCHSTGYCWCVTPDGRISTAAVHKTPRC---- 153  
OY 100 AALLTCOMQAIIVNPGW-----CGPPSCADGSFDEVQCCASNGECY 143  
DB 154 -----PGSVNEKLPRREGTKTDDAAALALETPQGDDEDIASRYPTLM 197  
OY 144 CVDKKELEGTROGGRPTCERHLSCECEARIKAHNSLRVEMFVPECLEDGSYNPVOCW 203  
DB 198 TEQVKSRONTKNNKPNVSSCDQEHQSALEAKQPKNDN-----VVIPECAHGILKRYQCH 252  
OY 204 PSTGYCWCY-DEGGVAVPGSDVFRKRPYC 231  
DB 253 PSTGYCWCYLVDTGRPLPSTIRYEQPC 281

RESULT 9  
O04901 PRELIMINARY: PRT: 1161 AA.  
ID O04901.  
AC O04901.  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE NIDOGEN PRECURSOR (ENTACTIN).  
OS Halocynthia roretzi (Sea squirt).  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
OC Stolidobranchia; Pyuridae; Halocynthia.  
OX NCBI\_TaxID=7729;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=MUSCLE;  
RX MEDLINE=93238676; PubMed=8477687;  
RA Nakae H., Sugano M., Ishimori Y., Endo T., Obinata T.;  
RT "Ascidian entactin/nidogen. Implication of evolution by shuffling two  
RT kinds of cysteine-rich motifs.";  
RL Eur. J. Biochem. 213:11-19(1993).  
CC -1- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN  
CC BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.  
CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-  
CC EXTRACELLULAR MATRIX INTERACTIONS.  
CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.  
CC -1- PTM: POSSESSES BOTH N- AND O-LINKED OLIGOSACCHARIDES.  
CC -1- SIMILARITY: TO EPIDERMAL GROWTH FACTOR PRECURSOR AND LOW DENSITY  
CC LIPOPROTEIN RECEPTOR.  
CC EMBL: D14038; BAA03127.1; -.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR000033; ldl\_receptor\_rep.  
DR InterPro: IPR003886; Nidogen\_ext.  
DR InterPro: IPR00716; thyroglobulin\_1.  
DR InterPro: IPR000834; zn\_carboxypept.  
DR Pfam: PF00058; ldl\_recept\_b; 4.  
DR Pfam: PF00086; thyroglobulin\_1; 3.  
DR SMART: SM00181; EGF; 1.  
DR SMART: SM00135; Lf; 4.  
DR SMART: SM00539; NIDO; 1.  
DR SMART: SM00211; TY; 3.  
DR PROSITE: PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_1.  
DR PROSITE: PS01186; EGF\_2; 1.  
DR PROSITE: PS00484; THYROGLOBULIN\_1; UNKNOWN\_2.  
KW Basement membrane; Calcium-binding; Cell adhesion; EGF-like domain;  
KW Extracellular matrix; Glycoprotein; Repeat; Signal; Sulfatation.  
FT SIGNAL 1 20  
FT CHAIN 21 1161  
FT DOMAIN 21 549  
FT DOMAIN 550 816  
FT DOMAIN 559 573  
FT DOMAIN 817 1161  
IT I (LARGER GLOBULAR DOMAIN) (BY  
SIMILARITY).  
II (CYSTEINE-RICH) (BY SIMILARITY).  
CONTAINS 5 X 3 AA TANDEM REPEATS OF R-V-  
P.  
III (SMALLER GLOBULAR DOMAIN) (BY

FT	REPEAT	276	306	SIMILARITY).
FT	REPEAT	550	605	EGF-LIKE.
FT	REPEAT	559	561	EGF-LIKE.
FT	REPEAT	562	564	1.
FT	REPEAT	565	567	2.
FT	REPEAT	568	570	3.
FT	REPEAT	571	573	4.
FT	REPEAT	606	664	5.
FT	REPEAT	665	704	EGF-LIKE.
FT	REPEAT	705	749	EGF-LIKE.
FT	REPEAT	1128	1161	EGF-LIKE.
FT	CARBOHYD	107	137	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	334	304	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	360	360	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	484	484	N-LINKED (GLCNAC. . .)
FT	SEQUENCE	1161 AA:	128471 MW:	(POTENTIAL).

Query Match	18.5%;	Score 238.5;	DB 5;	Length 1161;
Best Local Similarity	32.3%;	Pred. No. 1.1e-16;		
Matches 72;	Conservative 31;	Mismatches 83;	Indels 37;	Gaps 13

```

OY 28 TSTESLTKCO-QLOASNSGLIG-----HYVROCKETGEFEKOCWGSNGCYACED 80
    || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 555 TTFE-PMVCEERIEQEAERMDLCKELADELFFPTCTNLGTFAPROCTTIDSVCMCYDOD 655
OY 81 -GKEILGTRKINGSP-----DCSRKK--FALTLICOMMOALIVVPWCGP-----PS 123
    || || || || || || || || || || || || || || || || || || || || ||
Db 654 SGVEIDGR--KSLDREGYGLDCSDVRRPPPSQTPCEEDRTIAENARRRAGSEANSIMPR 711
OY 124 CKADSGDEVOCCASNGECIYVDKKKKELESTROOGRPTCRHLSECEEARIKHNSL- 183
    || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 712 CMESGAYEAPVQ-QLPDGTGFMVCVDDEGEIISRSREE--PRC---LTSCQLAAYVAESNNIF 765
OY 183 -RVEMFVPECLDEGSYNVQCMPSGTGCMVCYDEGVAPGSDV 224
    || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 766 KDTSLFVVOCKPDEGEYLPKROCVAS--VCEVCADRNGLELNSNRV 806

```

RESULT	10
Q98UD0	
Q98UD0	PRELIMINARY; PRT; 88 AA.
AC	Q98UD0
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	OVARIAN CYSTEINE PROTEASE INHIBITOR.
GN	OCPI.
OS	Salvelinus fontinalis (Brook trout) (Brook char).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC	Protacanthopterygii; Salmoniformes; Salmonidae; Salvelinus.
OX	NCBI_TaxId=8038;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Bohe J., Goetz F.W.;
RT	"Identification of a cysteine protease inhibitor mRNA in the ovulatory
RT	and postovulatory brook trout ovary."
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBSJ databases.
DR	EMBL; AF223387; AAK00216.1; -.
KW	Protease.
SO	SEQUENCE 88 AA; 9171 MW; F14E73AC2884F9E8 CRC64;

Query Match	13.8%;	Score 177.5;	DB 13;	Length 88;
Best Local Similarity	38.2%;	Pred. No. 2.1e-11;		
Matches	29;	Conservative	18;	Mismatches 26;
			Indels	3;
			Gaps	1

```
Oy      16 VMIVWFACAITSTEASL--TKQOOLQASANSGLIGTYVPCKETGEFEEEKOCMSGTG 72
       :: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      3 ILTILLVSTAFLGDATIRPKTPCEARARDATHGPICAYIPTCDAAGQYTPKQCWSAG 62
```

```
QY      73 YCWCVDDEDGKEILGK 88
      ||||| |:| ||:
Db      63 YCWCVTSTGQKIQTGTE 78
```

RESULT	11
Q9V5F6	
ID	Q9V5F6
PRELIMINARY;	
PRT;	613 AA

DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE CG2264 PROTEIN.  
GN CG2264.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Empidoidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Branton R.C., Rogers Y.-H.C., Blazer R.G., Chang M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abell J.F., Aghayani A., An H.-J., Andrews-Plankow C., Baldwin D.,  
RA Bailew R.M., Basu A., Baxendale J., Bayraktiroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,  
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Haverson D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
RA Jajjala M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodire C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Lin X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mekullov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Sliden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stappleton M., Strong R., Sun E.,  
RA Svitskys R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weisenbach J.,  
RA Williams S.W., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu G., Zhu H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.C.,  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -1- SIMILARITY: TO EF-HAND FAMILY.  
DR EMBL: AE003831; AAF58954.1; -.  
DR HSSP: P80424; IAN1.  
DR FlyBase: FBgn0033478; CG2264.  
DR InterPro: IPR002048; EF-hand.  
DR InterPro: IPR002350; kazal.  
DR InterPro: IPR00716; Rhyroglobulin\_1.  
DR Pfam: PF00036; ehand; 2.  
DR Pfam: PF00050; kazal; 1.  
DR Pfam: PF00086; rhyroglobulin\_1; 2.  
DR SMART: SM00280; KAZAL; 1.  
DR SMART: SM00211; TY; 2.



Db 1119 TGHMSPVQCLGKOPQPMDRHTEIVSRAFASEAPASAGEAPGVWCADKKGAPLKGTLTR 1178  
 QY 226 FKRPCT 231  
 Db 1179 ESEPIC 1184

# RESULT 13

Q9HCY2 PRELIMINARY; PRT; 126 AA.

AC 09HCY2;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE D421D16.1 (NOVEL THYROGLOBULIN TYPE-1 REPEAT CONTAINING PROTEIN)  
 DE (FRAGMENT).  
 GN D421D16.1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stycamore N.;  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL109940; CAC03594.1; -;  
 DR InterPro; IPR002350; kazal.  
 DR InterPro; IPR000716; Thyroglobulin\_1.  
 DR Pfam; PF00086; kazal.1.  
 DR Pfam; PF00086; thyroglobulin\_1.  
 DR SMART; SM00280; KAZAL.1.  
 DR SMART; SM00211; TY.1.  
 DR PROSITE; PS00484; THYROGLOBULIN\_1.  
 FT NON\_TER 1 1  
 FT NON\_TER 126 126  
 SQ SEQUENCE 126 AA; 14281 MW; D7076BEF2B89ABDA CRC64;

## Query Match

Best Local Similarity 25.6%; Score 167.5; DB 4; Length 126;  
 Matches 43; Conservative 17; Mismatches 49; Indels 59; Gaps 4;

QY 77 VDEGKEILGKIRGSPDCSRRAALTLQMMQAIIVNPGWGPSPCKADG-----S 129  
 Db 4 VDQDKDCSLDCAGSPQ-----KPLCAGDGRPLSRCE 37  
 QY 130 FDEYQC-----CASNECCVADKKKKELEGTROQGRPTGERHLSCEEARIRAHNSLR 183  
 Db 38 FORAKCDPQLEIAYRGNCRDVSRCVARERYTQEQARKERO-----S 78  
 QY 184 VEMFVPCLEBDGSYNPQCPSTGYCWCVDGEGVYKPGSDVRRFRPTC 231  
 Db 79 -QVFIPECNDDGITYSOVQCHSYTGICWCVTNPGRIISGTIVAAKTPRC 125

# RESULT 14

Q9XRE4 PRELIMINARY; PRT; 234 AA.

AC 09XRE4;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE MHC CLASS II-ASSOCIATED INVARIANT CHAIN.  
 OS Brachydanio rerio (zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Rasbortinae; Danio.  
 OX NCBI\_Taxid=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-EX; TISSUE-SPLEEN;  
 RA Natarajan K., Castellano F., Germain R.N., Margulies D.H.;  
 RT "The zebrafish invariant chain promotes the transport and maturation

RT of mouse MHC-II molecules.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF114830; AAD24542.1; -;  
 DR InterPro; IPR000716; Thyroglobulin\_1.  
 DR Pfam; PF00086; thyroglobulin\_1.  
 DR PROSITE; PS00484; THYROGLOBULIN\_1; UNKNOWN.1.  
 DR SMART; SM00211; TY.1.  
 SQ SEQUENCE 234 AA; 25916 MW; 76C5917D7DFA6362 CRC64;

Query Match 13.0%; Score 167; DB 7; Length 234;  
 Best Local Similarity 50.8%; Pred. No. 8e-10;  
 Matches 31; Conservative 5; Mismatches 23; Indels 2; Gaps 1;

QY 35 TKGQQLQASANGSLGTYTPCKEKGFEFEKQCMSTGCMCVDDGKEILGKIRGSPD 94  
 Db 176 TRCK--LSESEYRVRGFEKFPACDERGNYPMQCMHSTGCMCVTDGTEIGSTRIRGRPQ 233  
 QY 95 C 95  
 Db 234 C 234

# RESULT 15

Q9PUT2 PRELIMINARY; PRT; 237 AA.

AC 09PUT2;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE INVARIANT CHAIN-LIKE PROTEIN 1.  
 GN ICLP-1.  
 OS Brachydanio rerio (zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Rasbortinae; Danio.  
 OX NCBI\_Taxid=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-20009532; PubMed=10541812;  
 RT Yoder J.A., Haire R.N., Litman G.W.;  
 RT "Cloning of two zebrafish cDNAs that share domains with the MHC class II-associated invariant chain.";  
 RT Immunogenetics 50:84-88(1999).  
 RL EMBL; AF148214; AAF01568.1; -;  
 DR InterPro; IPR000716; Thyroglobulin\_1.  
 DR Pfam; PF00086; thyroglobulin\_1.  
 DR PROSITE; PS00484; THYROGLOBULIN\_1; UNKNOWN.1.  
 DR SMART; SM00211; TY.1.  
 SQ SEQUENCE 237 AA; 26231 MW; E56D326C5917D7D CRC64;

Query Match 13.0%; Score 167; DB 13; Length 237;  
 Best Local Similarity 50.8%; Pred. No. 8.1e-10;  
 Matches 31; Conservative 5; Mismatches 23; Indels 2; Gaps 1;

QY 35 TKGQQLQASANGSLGTYTPCKEKGFEFEKQCMSTGCMCVDDGKEILGKIRGSPD 94  
 Db 176 TRCK--LSESEYRVRGFEKFPACDERGNYPMQCMHSTGCMCVTDGTEIGSTRIRGRPQ 233  
 QY 95 C 95  
 Db 234 C 234

Search completed: March 8, 2002, 09:25:59  
 Job time: 117 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2002, 09:21:52 ; Search time 17.1 Seconds  
(without alignments)  
1029.025 Million cell updates/sec

Title: US-09-445-480a-2

Sequence score: 1287  
1 MALSONQAKFKSGFVYIMV.....VDEGVKVPESDVRKRPPTC 231

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	295	22.9	2769	1	UIBO
2	279.5	21.7	243	2	thyroglobulin prec
3	279	21.7	212	2	thyroglobulin 2 pr
4	271	21.1	2767	1	thyroglobulin prec
5	245.5	19.1	1376	2	thyroglobulin prec
6	238.5	18.5	1161	1	osteonidogen - hum
7	165	12.8	2225	2	nidogen precursor
8	159	12.4	439	2	hypothetical prote
9	158.5	12.3	2150	2	testican - human
10	154	12.0	1245	1	hypothetical prote
11	153	11.9	280	2	nidogen precursor
12	153	11.9	1247	1	class II histocomp
13	150	11.7	71	2	nidogen precursor
14	135	10.5	279	1	protein 41 - human
15	132.5	10.3	258	1	class II histocomp
16	129.5	10.1	254	2	insulin-like growt
17	129	10.0	254	2	insulin-like growt
18	128.5	10.0	254	2	insulin-like growt
19	124.5	9.7	237	2	insulin-like growt
20	122.5	9.5	258	1	insulin-like growt
21	122.5	9.5	271	2	insulin-like growt
22	122.5	9.5	272	2	insulin-like growt
23	120.5	9.4	271	2	insulin-like growt
24	118.5	9.2	157	2	thyroglobulin - ra
25	118.5	9.2	271	2	thyroglobulin - ra
26	114.5	8.9	1167	2	insulin-like growt
27	114	8.9	263	1	hypothetical prote
28	114	8.9	1786	1	insulin-like growt
29	113.5	8.8	240	2	lamin beta-1 cha
					insulin-like growt

#### ALIGNMENTS

```

30      112      8.7      369      2      T29836      hypothetical prote
31      112      8.7      1786      1      MMHUB1      laminin beta-1 cha
32      111.5      8.7      949      2      T24294      hypothetical prote
33      110      8.5      266      1      A35037      insulin-like growt
34      110      8.5      291      1      JN0064      insulin-like growt
35      109.5      8.5      259      1      IOH01      insulin-like growt
36      109.5      8.5      4544      1      S02392      alpha-2-macroglobu
37      109      8.5      291      2      I48602      insulin-like growt
38      108.5      8.4      252      1      A36748      insulin-like growt
39      108      8.4      272      1      A36082      insulin-like growt
40      108      8.4      1620      2      T27283      Spot 3 protein and
41      107.5      8.4      623      2      T47542      insulin-like growt
42      107      8.3      226      2      JN0464      insulin-like growt
43      107      8.3      238      2      I48605      insulin-like growt
44      107      8.3      291      1      IOH03      insulin-like growt
45      107      8.3      305      2      I48601      insulin-like growt

RESULT 1
UIBO
C:Species: Bos primigenius taurus (cattle)
C:Date: 28-May-1986 #sequence.revision 28-May-1986 #text_change 16-Jun-2000
C:Accession: A01533; S10078; S05566; A32735; S39431; A32740; S29734; I45981
R:Mercken, L.; Simons, M.J.; Swillens, S.; Masser, M.; Vassart, G.
A:Title: Primary structure of bovine thyroglobulin deduced from the sequence of 1
A:Reference number: A01533; MUID:85296288
A:Accession: A01533
A:Molecule type: mRNA
A:Residues: 1-2769 <MER>
A:Cross-references: GB:X02815; GB:J00040; NID:9774; PIDN:CA26584.1; PID:9775
R:Hansen, C.; Gerard, C.; Vassart, G.; Stordeur, P.; Christophe, D.
Eur. J. Biochem. 178, 387-393, 1988
A:Title: Thyroid-specific and cAMP-dependent hypersensitive regions in thyroglobu
A:Reference number: S10078; MUID:89091173
A:Accession: S10078
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-20 <HAN>
A:Cross-references: EMBL:X14324; NID:9788; PIDN:CA32504.1; PID:9789
R:Parma, J.; Christophe, D.; Pohl, V.; Vassart, G.
J. Mol. Biol. 196, 769-779, 1987
A:Title: Structural organization of the 5' region of the thyroglobulin gene. Evic
A:Reference number: S03422; MUID:88062712
A:Accession: S05566
A:Molecule type: DNA
A:Residues: 1002-1205, 'R', 1207-1209 <PAR>
A:Cross-references: EMBL:X06071; NID:9778; PIDN:CA29457.1; PID:92204111
R:Rawlitch, A.B.; Lilwer, M.R.; Gregg, J.; Turner, C.D.; Rouse, J.B.; Hamilton, J.
Biochem. Biophys. Res. Commun. 118, 423-429, 1984
A:Title: The isolation of identical thyroxine containing amino acid sequences fro
A:Reference number: A32735; MUID:84153804
A:Accession: A32735
A:Molecule type: protein
A:Residues: 20-38 <RAW>
R:Genille, F.; Salvatore, G.
Eur. J. Biochem. 218, 603-621, 1993
A:Title: Preferential sites of proteolytic cleavage of bovine, human and rat thyr
A:Reference number: S39431; MUID:94094855
A:Accession: S39431
A:Molecule type: protein
A:Residues: 20-27;259-265;521-526, 'X', 528;539-560;1161-1168, 'XX', 1171-1172;1203-1
R:Rawlitch, A.B.; Mercken, L.; Hamilton, J.W.; Vassart, G.
Biochem. Biophys. Res. Commun. 119, 335-342, 1984
A:Title: The structure of a naturally occurring 10K polypeptide derived from the
A:Reference number: A32740; MUID:84153919
A:Accession: A32740
A:Molecule type: protein
A:Residues: 20-38;66-78;95-99 <RA2>

```

R.Rawitch, A.B.; Pollock, H.G.; Yang, S.X.  
Arch. Biochem. Biophys. 300, 271-279, 1993  
A>Title: Thyroglobulin glycosylation: location and nature of the N-linked oligosaccharid  
A:Reference number: S29734; MUID:93143326  
A:Accession: S29734  
A:Molecule type: protein  
A:Residues: 106-109, 'X', 111-117, 475-477, 'N', 479-482, 'X', 484-485, 492-494, 'X', 496-502, 839-  
252-254, 2272-2281 <RA3>  
R:Ricketts, M.H.; Simons, M.  
Proc. Natl. Acad. Sci. U.S.A. 84, 3181-3184, 1987  
A>Title: A nonsense mutation causes hereditary goitre in the Afrikaner cattle and unmas  
A:Reference number: 145981; MUID:87204101  
A:Accession: 145981  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 712-723 <R1C>  
A:Cross-references: GB:M16448; NID:9163743; PIDN:AAA30777.1; PID:g552340  
C:Comment: The thyroglobulin molecule is a dimeric glycoprotein specific to the thyroid  
C:Genetics:  
A:Introns: 1047/1, 1073/1, 1110/3, 1145/1  
A>Note: the list of introns may be incomplete  
C:Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat homol  
C:Keywords: duplication; glycoprotein; homodimer; iodine; thyroid gland; thyroid hormone  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-2769/Product: thyroglobulin #status predicted <TGL>  
F:34-92/Domain: thyroglobulin type I repeat homology <TH01>  
F:96-160/Domain: thyroglobulin type I repeat homology <TH02>  
F:164-297/Domain: thyroglobulin type I repeat homology #status atypical <TH03>  
F:301-358/Domain: thyroglobulin type I repeat homology <TH04>  
F:607-657/Domain: thyroglobulin type I repeat homology <TH05>  
F:661-725/Domain: thyroglobulin type I repeat homology <TH06>  
F:729-921/Domain: thyroglobulin type I repeat homology #status atypical <TH07>  
F:925-1073/Domain: thyroglobulin type I repeat homology #status atypical <TH08>  
F:1077-1145/Domain: thyroglobulin type I repeat homology <TH09>  
F:1146-1210/Domain: thyroglobulin type I repeat homology <TH10>  
F:1458-1471/Region: type II repeat  
F:1472-1488/Region: type II repeat  
F:1489-1505/Region: type II repeat  
F:1605-1725/Region: type III repeat  
F:1726-1893/Region: type III repeat  
F:1894-1996/Region: type III repeat  
F:1997-2130/Region: type III repeat  
F:2131-2188/Region: type III repeat  
F:2229-2777/Domain: cholinesterase homology <ACE>  
F:24/Modified site: thyroxine (Tyr) #status experimental  
F:110,483,495,853,947,1140,1365,1776,1870,2014,2123,2251,2296/Binding site: carbohydrate  
F:198/Binding site: carbohydrate (Asn) (covalent) #status absent  
F:2574/Modified site: thyroxine (Tyr) #status predicted  
F:2588/Modified site: thyroxine (Tyr) #status predicted  
F:2767/Modified site: triiodothyronine (Tyr) #status predicted

Query Match 22.9%; Score 295; DB 1; Length 2769;  
Best Local Similarity 33.5%; Pred. No. 3,1e-16;  
Matches 69; Conservative 21; Mismatches 64; Indels 52; Gaps 8;  
QY 52 YVPOCKETEPEFEKOCWGSGTCVDEGKELIGTKIRGSPDCSRRAATLTCOMQAI 111  
DB 48 YVPOCAEDSGFOTVOCGKAGASCVCADADGRRVPSRQGRP-----AACISFCOLQKQ 102  
QY 112 I-----VNVPGWCGPSCAGSFEVQCCASNGECYCVKKGKELGTRQGRPT-CER 165  
DB 103 ILLSYNSTATSYLPQCDSGYSVPCDLRRQOCWCVDAGMEVYGRQGRAPR 162  
QY 166 HLESECE-ERIKRHSNLSLVEFMVPECELEDSYNPQC----- 202  
DB 163 ---SCEIRNRRLHGVGDSP---PQCSPDGAFRPVQCKLVNTTDMIFDLVHSYRPPD 216  
QY 203 -----WPS-TGYCWCVDEGG 216  
DB 217 AFTFSSFRSFRPEVSGCYCCADSDG 242

RESULT 2  
A36124  
Thyroglobulin 2 precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 14-Dec-1990 #sequence\_revision 14-Dec-1990 #text\_change 11-Apr-1997  
C:Accession: A36124  
R:Graves, P.N.; Davies, T.F.  
Mol. Endocrinol. 4, 155-161, 1990  
A>Title: A second thyroglobulin messenger RNA species (rTg-2) in rat thyrocytes.  
A:Reference number: A36124; MUID:90220642  
A:Accession: A36124  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-243 <GRA>  
A:Cross-references: GB:M35965  
C:Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat  
C:Keywords: dimer; glycoprotein; iodine; thyroid gland; thyroid hormone biosynthes  
F:35-93/Domain: thyroglobulin type I repeat homology <THY1>  
F:97-161/Domain: thyroglobulin type I repeat homology <THY2>  
F:25/Modified site: thyroxine (Tyr) #status predicted

Query Match 21.7%; Score 279.5; DB 2; Length 243;  
Best Local Similarity 34.0%; Pred. No. 6,7e-16;  
Matches 69; Conservative 29; Mismatches 78; Indels 27; Gaps 8;  
QY 16 VNIWVLFICATITSEASTKQ-----OLQASANSGLITVYPOCKETEPEFEKQ 66  
DB 4 LVIMWSTLSSVCLVANAIFETVDAOPLRCEIQRKAEALFKODEYVPOCSDESGFOTVQ 63  
QY 67 CWGSTGYCWCVDEGKELIGTKIRGSPDCSRRAATLTCOM-MQAI1-----VNVPGWCGP 121  
DB 64 QNDGSCWCVCVDSGTEVPSRQLRP-----TACLSFQJHQRILSYNSTDALYL 118  
QY 122 PSCKAGSDEVQCCASNGECYCVKKGKELGTRQGRPT-CERHLECE-ERIKRHS 179  
DB 119 PQCDGSGNAPVQCDLQVQWCVCPESEMEVYGRQGRAPR---SCEIRNRRLHG 175  
QY 180 NSLRVEMFPECELEDSYNPQC 202  
DB 176 VG---DKSPPOCDADGRRVPSRQGRP 195

RESULT 3  
I46176  
Thyroglobulin precursor - goat (fragment)  
C:Species: Capra aegagrus hircus (domestic goat)  
C:Date: 14-Feb-1997 #sequence\_revision 14-Feb-1997 #text\_change 16-Jul-1999  
C:Accession: I46176  
R:Van Ommen, G.J.; Sterk, A.; Mercken, L.O.; Arnborg, A.C.; Baas, F.; De Vijlder, A.  
Biochimie 71, 211-221, 1989  
A>Title: Studies on the structures of the normal and abnormal goat thyroglobulin ge  
A:Reference number: I46176; MUID:89207610  
A:Accession: I46176  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-212 <VAN>  
A:Cross-references: EMBL:X14997; NID:9989; PIDN:CA33105.1; PID:g9990  
C:Genetics:  
A:Gene: Tg  
C:Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat  
F:34-92/Domain: thyroglobulin type I repeat homology <THY1>  
F:96-160/Domain: thyroglobulin type I repeat homology <THY2>

Query Match 21.7%; Score 279; DB 2; Length 212;  
Best Local Similarity 39.2%; Pred. No. 6,6e-16;  
Matches 62; Conservative 18; Mismatches 60; Indels 18; Gaps 6;  
QY 52 YVPOCKETEPEFEKOCWGSGTCVDEGKELIGTKIRGSPDCSRRAATLTCOMQAI 111  
DB 48 YVPOCAEDSGFOTVOCGKAGASCVCADADGRRVPSRQGRP-----AACISFCOLQKQ 102

```

Oy      112  I---VAVPEWCGPSPCKADGSDFEVQCCASNGECVCVKKKECTGPGQGRPT-CER 165
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      103  ILLSYINSTASYLPPCCQDSGDSPVQCILRRRQCVCVAEGMEYVTRQGRPAQCPR 162
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      166  HLSECE-EARIKAHNSLRLVEMFVEPCLEDGSGYNPVQC 202
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      163  ---SCELRNRRLHGVDRSP---PQCSPDGAFRPPVQC 194
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT      4
UIHU
thyroglobulin precursor, major splice form - human
C.Species: Homo sapiens (man)
C.Date: 28-Feb-1986 #sequence.revision 05-Nov-1999 #text.change 16-Jun-2000
C.Accession: A59110; S00014; A01532; S03422; I08343; I57669; S02266; S39432; S66241; S62
R.Malthiery, Y.
submitted to the EMBL Data Library, April 1988
A.Description: Human mRNA for thyroglobulin.
A.Reference number: A59110
A.Accession: A59110
A>Status: translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-2767 <MAL>
A.Cross-references: GB:X05615; NID:g37173; PIDN:CAA29104.1; PID:g37174
A.Note: revision to S00014
R.Malthiery, Y.; Lissitzky, S.
Eur. J. Biochem. 165, 491-498, 1987
A>Title: Primary structure of human thyroglobulin deduced from the sequence of its 8448-
A.Reference number: S00014; MUID:87246630
A.Accession: S00014
A.Molecule type: mRNA
A.Residues: 1-1041, 'Y', 1043-1057, 'T', 1059-2767 <MAL2>
A.Cross-references: GB:X05615; NID:g37173
A.Note: this sequence is revised in A59110
R.Malthiery, Y.; Lissitzky, S.
Eur. J. Biochem. 147, 53-58, 1985
A>Title: Sequence of the 5'-end quarter of the human-thyroglobulin messenger ribonucleic
A.Reference number: A01532; MUID:85127024
A.Accession: A01532
A.Molecule type: mRNA
A.Residues: 1-730 <MAN>
A.Cross-references: GB:X02154; NID:g37175; PIDN:CAA26089.1; PID:g135349
A.Note: the translated sequence in Genbank entry HSHYHR5, release 111.0, (PIDN:CAA26089
n 1-Met
R.Parma, J.; Christophe, D.; Pohl, V.; Vassart, G.
J. Mol. Biol. 196, 769-779, 1987
A>Title: Structural organization of the 5' region of the thyroglobulin gene. Evidence for
A.Reference number: S03422; MUID:88062712
A.Accession: S03422
A.Molecule type: DNA
A.Residues: 1-134, 'Q', 136-415; 640-652, 'G', 654-733, 'A', 735-737; 880-983, 'DR', 985-999 <PAR>
A.Cross-references: EMBL:X06059; NID:g37145; PIDN:CAA29454.1; PID:g1353884; EMBL:X06067;
R.Christophe, D.; Cabrier, B.; Baccolla, A.; Targovnik, H.; Pohl, V.; Vassart, G.
Nucleic Acids Res. 13, 5127-5144, 1985
A>Title: An unusually long poly(purine)-poly(pyrimidine) sequence is located upstream from
A.Reference number: I08343; MUID:85269632
A.Accession: I08343
A.Status: translation not shown; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-22, 'GKF' <CHR>
A.Cross-references: EMBL:X02749; NID:g37162; PIDN:CAA26527.1; PID:g758106
R.Targovnik, H.M.; Cochaux, P.; Corach, D.; Vassart, G.
Mol. Cell. Endocrinol. 84, R23-R26, 1992
A>Title: Identification of a minor tg mRNA transcript in RNA from normal and goitrous th
A.Reference number: I57669; MUID:92347597
A.Accession: I57669
A.Status: translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1503-1508, 'L', 1567-1601 <REM>
A.Cross-references: GB:S40807; NID:g252170; PIDN:AA82665.1; PID:g252171
A.Note: this sequence fragment represents a minor splice form
R.Mairiq, C.; Lejeune, P.; Vennot, N.; Vinec, L.
FEBS Lett. 242, 414-418, 1989

```

A:Title: Hormone synthesis in human thyroglobulin: possible cleavage of the poly  
A:Reference number: S02266; MUID:8912111  
A:Contents: evidence for Tyr-149 as ring donor for thyroxine formation  
A:Accession: S02266  
A:Molecule type: protein  
A:Residues: 101-109;114-121;126-131;143-149 <MAR>  
A:Note: only the first peptide was sequenced; others were isolated and their amin  
R:Gentile, F.; Salvatore, G.  
Eur. J. Biochem. 218, 603-621, 1993  
A:Title: Preferential sites of proteolytic cleavage of bovine, human and rat thy  
A:Reference number: S39431; MUID:94094855  
A:Accession: S39432  
A:Molecule type: protein  
A:Residues: 20-27;522-527;541-547;616-619, 'X', 621;1001-1005;1009-1011, 'X', 1013;14  
R:Xiao, S.; Pollock, H.G.; Taugog, A.; Rawitch, A.B.  
Arch. Biochem. Biophys. 320, 96-105, 1995  
A:Title: Characterization of homomeric sites in an N-terminal, cyanogen bromid  
A:Reference number: S66241; MUID:95314327  
A:Accession: S66241  
A:Molecule type: protein  
A:Residues: 'D', 21-23;45-47;147-148;150-152 <XIA>  
R:Yang, S.X.; Pollock, H.G.; Rawitch, A.B.  
Arch. Biochem. Biophys. 327, 61-70, 1996  
A:Title: Glycosylation in human thyroglobulin: location of the N-linked oligosac  
A:Reference number: S62778; MUID:96201348  
A:Accession: S62778  
A:Molecule type: protein  
A:Residues: 69-84;196-210, 'T', 212-213;476-492;523-539;741-770;811-848, 850-853;938  
A:1783, 'D', 1785, 'R', 1999-2017;2241-2258;2270-2285, 'P', 2287-2307;2578-2584 <YAN>  
R:Reit, T.; Cochaux, P.; Targovnik, H.M.; Suzuki, M.; Shimoda, S.; Perret, J.; V  
J. Clin. Invest. 88, 1901-1905, 1991  
A:Title: A 3' splice site mutation in the thyroglobulin gene responsible for cong  
A:Reference number: I55655; MUID:92091498  
A:Accession: I55655  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 90-91,160-161 <IEI>  
A:Cross-References: GB:S71821; NID:9240935; PIDN:AA820655.1; PID:9240936  
A:Note: mutant splice form  
C:Comment: The thyroglobulin molecule is produced in the thyroid gland and is the  
C:Genetics:  
A:Gene: GDB:7G  
A:Cross-References: GDB:120434; OMIM:188450  
A:Map position: 89q24.2-8q24.3  
A:Introns: 23/1; 59/2; 92/1; 160/1; 213/2; 249/1; 297/1; 359/1; 726/1; 921/1; 100  
C:Complex: homodimer  
C:Function:  
A:Description: precursor of thyroid hormones thyroxine and triiodothyronine; iod  
C:Superfamily: thyroglobulin: cholinesterase homology; thyroglobulin type I repe  
C:Keywords: alternative splicing; duplication; glycoprotein; homodimer; iodine; t  
F:1-19/Domain: signal sequence #status predicted <SIC>  
F:20-2167/Product: thyroglobulin #status predicted <MAT>  
F:34-92/Domain: thyroglobulin type I repeat homology <THY1>  
F:56-160/Domain: thyroglobulin type I repeat homology <THY2>  
F:164-297/Domain: thyroglobulin type I repeat homology <THY3>  
F:301-358/Domain: thyroglobulin type I repeat homology <THY5>  
F:608-658/Domain: thyroglobulin type I repeat homology <THY6>  
F:662-726/Domain: thyroglobulin type I repeat homology #status atypical <THY7>  
F:730-921/Domain: thyroglobulin type I repeat homology #status atypical <THY8>  
F:925-1072/Domain: thyroglobulin type I repeat homology #status atypical <THY8>  
F:1076-1144/Domain: thyroglobulin type I repeat homology <THY9>  
F:1148-1209/Domain: thyroglobulin type I repeat homology <TH10>  
F:1455-1468/Region: type II repeat  
F:1469-1485/Region: type II repeat  
F:1486-1502/Region: type II repeat  
F:1602-1722/Region: type IIIa repeat  
F:1723-1891/Region: type IIIb repeat  
F:1892-1994/Region: type IIIa repeat  
F:1995-2128/Region: type IIIb repeat  
F:2129-2186/Region: type IIIa repeat  
F:2227-2725/Domain: cholinesterase homology <ACE>  
F:24, 2572, 2576/Modified site: thyroxine (Tyr) #status predicted  
F: 76, 158, 184, 529, 748, 816, 947, 1219, 1348, 1364, 1715, 1773, 2012, 2249, 2294, 2581/Binding









submitted to the EMBL Data Library, March 1988

A:Reference number: S03099

A:Accession: S03099

A:Molecule type: mRNA

A:Residues: 1-58 <STO>

A:Cross-references: EMBL:X07129; NID:952637; PIDN:CA30141.1; PID:952638

R:Zhu, L.; Jones, P.P.

Nucleic Acids Res. 17, 447-448, 1989

A:Title: Complete sequence of the murine invariant chain (II) gene.

A:Reference number: S09284; MUID:89098406

A:Accession: S09284

A:Molecule type: DNA

A:Residues: 1-9, 'M', 11-112, 'Q', 114-228, 'STG', 233-279 <ZHU>

R:Miller, J.; Hatch, J.A.; Simons, S.; Cullen, S.E.

Proc. Natl. Acad. Sci. U.S.A. 85, 1359-1363, 1988

A:Title: Identification of the glycosaminoglycan-attachment site of mouse invariant-chain

A:Reference number: A28956; MUID:88144436

A:Accession: A28956

A:Molecule type: mRNA

A:Residues: 261-270 <MTL>

A:Note: proof by site-directed mutagenesis that the sole chondroitin sulfate glycosaminoglycan

C/enetics:

A:introns: 26/2; 83/1; 109/3; 130/3; 161/3; 192/1; 256/1; 277/1

C:superfamily: class II histocompatibility antigen-associated gamma chain; thyroglobulin

C:keywords: alternative splicing; glycoprotein; transmembrane protein

F:1-279/Product: class II histocompatibility antigen-associated gamma chain II, 41K sp1

F:1-191,256-279/Product: class II histocompatibility antigen-associated gamma chain II,

F:1-29/Domain: intracellular #status predicted <CYT>

F:30-55/Domain: transmembrane #status predicted <TM>

F:56-279/Domain: extracellular #status predicted <EXT>

F:196-254/Domain: thyroglobulin type I repeat homology <THY1>

F:113,119,239,253/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:265/Binding site: carbohydrate (Ser) (covalent) #status experimental

Query Match 10.5%; Score 135; DB 1; Length 279;

Best Local Similarity 37.2%; Pred. No. 0.00034;

Matches 52; Conservative 11; Mismatches 35; Indels 8; Gaps 2;

QY 19 WLEIFACATY-----STEAS---LTKCOOLASANSGLIGYVPQCKETGEFEKQCGS 70

DB 170 WLFEEKSKNSLEKKPTKPEVLFKQCEVSHIPAVYPAFAFRKPCDENGNLYLQCHGR 229

QY 71 TGCCWCVDEDDGKEILGTRKSPDCS 96

DB 230 HCCYWCVCFPNGTEVPTKSRGRNCS 255

RESULT 15

A45403

insulin-like growth factor-binding protein 4 - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 03-Dec-1999

C:Accession: A45403

R:Moser, D.R.; Lowe, J.R.; W.L.; Dake, B.L.; Booth, B.A.; Boes, M.; Clemmons, D.R.; Bar, F

Mol. Endocrinol. 6, 1805-1814, 1992

A:Title: Endothelial cells express insulin-like growth factor-binding proteins 2 to 6.

A:Reference number: A45403; MUID:93125553

A:Accession: A45403

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-258 <MOS>

A:Cross-references: GB:S52770; NID:g263303; PIDN:AAB24873.1; PID:g263304

A:Experimental source: pulmonary artery endothelial cells

A:Note: sequence extracted from NCBI backbone (NCBIIN:122183, NCBIIP:122184)

C:superfamily: insulin-like growth factor binding protein 1; thyroglobulin type I repeat

F:174-249/Domain: thyroglobulin type I repeat homology <THY1>

Query Match 10.3%; Score 132.5; DB 2; Length 258;

Best Local Similarity 23.0%; Pred. No. 0.00081;

Matches 52; Conservative 33; Mismatches 70; Indels 71; Gaps 13;

QY 37 COQLQASANSGL-----IGYVPQCKETGEFEKQCGSTGYCWCVEDDGKEIL 85  
DB 44 CEELVREPGGCGCATCATLGKGMPCGYTPRCG-----SGLRCYPRGV-----EKPL 90  
QY 86 GTRKSPDCSRRRAALTLCOAMQALIVNPGMCPSPCKAG-----SPDEVCCASNG 140  
DB 91 HTLVHGGVCVMELEIAIDQESIQ-----PSDKDGDHPNNSFS--PCSAHDR 136  
QY 141 ECVYCDK-----KGELEGTROQGRP-----TCERNLSECEARIKAHNSLR 183  
DB 137 K--CLQKHIAKIRDRSTSGGKMKVIGAPREARVPQSCQSELHRALE-RLAASQSRTH 193  
QY 184 VEMF-----VPECLDGSINPYQCPSP-----TGCCWCVDEDDGKEIL 221  
DB 194 EDLYTIPINCDNRGNENHPRQCHPALDGGRGKCMCVDRKTVGLKPG 239

Search completed: March 8, 2002, 09:25:12  
Job time: 200 sec



Mon Mar 11 08:00:36 2002

us-09-445-480a-2.rpt

Page 9



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: March 8, 2002, 09:21:32 ; Search time 12.57 Seconds  
(without alignments)

413.545 Million cell updates/sec

Title: US-09-445-480A-2  
Perfect score: 1287  
Sequence: 1 MALSONQAKFSKGFVYMWV.....VDEGVKVPKSDVRFKRPCT 231

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	122.5	9.5	186	1 US-08-698-551-6	Sequence 6, Appl1
2	122.5	9.5	186	2 US-08-602-228-6	Sequence 6, Appl1
3	122.5	9.5	186	2 US-08-649-341A-6	Sequence 6, Appl1
4	122.5	9.5	186	2 US-08-494-440B-6	Sequence 6, Appl1
5	122.5	9.5	186	2 US-08-533-901B-6	Sequence 6, Appl1
6	122.5	9.5	186	2 US-08-839-032A-6	Sequence 6, Appl1
7	122.5	9.5	186	2 US-08-839-031A-6	Sequence 6, Appl1
8	122.5	9.5	186	5 PCT-US95-12724-6	Sequence 6, Appl1
9	122.5	9.5	237	6 5212074-7	Patent No. 5212074
10	122.5	9.5	272	1 US-08-698-551-8	Sequence 8, Appl1
11	122.5	9.5	272	2 US-08-602-228-8	Sequence 8, Appl1
12	122.5	9.5	272	2 US-08-649-341A-8	Sequence 8, Appl1
13	122.5	9.5	272	2 US-08-494-440B-8	Sequence 8, Appl1
14	122.5	9.5	272	2 US-08-533-901B-8	Sequence 8, Appl1
15	122.5	9.5	272	2 US-08-839-032A-8	Sequence 8, Appl1
16	122.5	9.5	272	2 US-08-839-031A-8	Sequence 8, Appl1
17	122.5	9.5	272	5 PCT-US92-01196-4	Sequence 4, Appl1
18	122.5	9.5	272	5 PCT-US95-12724-8	Sequence 8, Appl1
19	122.5	9.5	271	5 PCT-US92-01196-3	Sequence 3, Appl1
20	113.5	8.8	213	6 5212074-6	Patent No. 5212074
21	113.5	8.8	322	6 5212074-1	Patent No. 5212074
22	112	8.7	1196	1 US-08-144-121-4	Sequence 4, Appl1
23	112	8.7	1196	2 US-08-735-893-4	Patent No. 5212074
24	109.5	8.5	259	6 5212074-3	Sequence 52, Appl1
25	109.5	8.5	4544	1 US-08-469-486-52	Sequence 52, Appl1
26	109.5	8.5	4544	2 US-08-469-658-52	Sequence 52, Appl1
27	107	8.3	85	3 US-08-604-965E-9	Sequence 9, Appl1

28	107	8.3	264	1 US-08-482-271-3	Sequence 3, Appl1
29	107	8.3	264	1 US-08-482-271-4	Sequence 4, Appl1
30	107	8.3	264	2 US-08-854-811-45	Sequence 45, Appl1
31	107	8.3	264	3 US-09-080-120A-2	Sequence 2, Appl1
32	107	8.3	264	3 US-09-080-120A-4	Sequence 4, Appl1
33	107	8.3	264	5 PCT-US95-08925-2	Sequence 2, Appl1
34	107	8.3	264	5 PCT-US95-08925-4	Sequence 4, Appl1
35	107	8.3	264	5 PCT-US95-08925-7	Sequence 19, Appl1
36	107	8.3	264	1 US-08-468-847B-19	Sequence 7, Appl1
37	107	8.3	264	1 US-09-080-120A-7	Sequence 7, Appl1
38	107	8.3	264	5 PCT-US95-08925-7	Sequence 7, Appl1
39	106.5	8.3	292	6 5212074-5	Patent No. 5212074
40	104	8.1	810	2 US-08-820-170A-34	Sequence 34, Appl1
41	104	8.1	810	3 US-09-055-699-34	Sequence 34, Appl1
42	104	8.1	810	4 US-09-273-565-34	Sequence 34, Appl1
43	103.5	8.0	1358	1 US-08-404-665-4	Sequence 4, Appl1
44	103.5	8.0	1358	1 US-08-404-671-4	Sequence 4, Appl1
45	103.5	8.0	1358	1 US-08-404-781-4	Sequence 4, Appl1

## ALIGNMENTS

```
RESULT 1
US-08-698-551-6
: Sequence 6, Application US/08698551
: Patent No. 5712381
:
: GENERAL INFORMATION:
: APPLICANT: Lin, Lih-Ling
: APPLICANT: Chen, Jennifer H.
: APPLICANT: Schiavella, Andrea
: APPLICANT: Graham, James
: TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
: TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESS: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02140
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA: US/08/698,551
: APPLICATION NUMBER: US/08/698,551
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Brown, Scott A.
: REGISTRATION NUMBER: 32,724
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 186 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-698-551-6
```

Query Match 9.5%; Score 122.5; DB 1; Length 186;  
Best Local Similarity 35.1%; Pred. No. 3.4e-05;  
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;  
QY 151 ELECTNOGPTCEERHLSDEEAKRIKAHSNSLRVEFVPECELEDSYNPQCMPS---T 206  
I: :||| |||| :||| :||| | | | |



ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A, 32,724  
REGISTRATION NUMBER: G15232A  
REFERENCE/DOCKET NUMBER: G15232A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 186 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-494-440B-6

Query Match 9.5%; Score 122.5; DB 2; Length 186;  
Best Local Similarity 35.1%; Pred. No. 3.4e-05;  
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

OY 151 ELECTROOGRPTCEHLSCEERIKAHNSLRVEMFVPECLDGSYNPOQWPS---T 206  
DB 96 EMROESEOG--PCRRHM-EASLOELKASPRMVPRAVYLPNCDRKGFTKRKCKPSGRRR 152  
OY 207 GYCWCVDEGGVKKVPGSD 223  
DB 153 GICWCVDKYGKMKLPQME 169

RESULT 5  
US-08-533-901B-6

Sequence 6, Application US/08533901B  
Patent No. 5852173  
GENERAL INFORMATION:  
APPLICANT: Lin, Lih-Ling  
APPLICANT: Chen, Jennifer H.  
APPLICANT: Schievella, Andrea  
APPLICANT: Graham, James  
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND  
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/533,901B  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A,  
REGISTRATION NUMBER: 32,724  
REFERENCE/DOCKET NUMBER: G15232  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 186 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-533-901B-6

Query Match 9.5%; Score 122.5; DB 2; Length 186;  
Best Local Similarity 35.1%; Pred. No. 3.4e-05;  
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

OY 151 ELECTROOGRPTCEHLSCEERIKAHNSLRVEMFVPECLDGSYNPOQWPS---T 206  
DB 96 EMROESEOG--PCRRHM-EASLOELKASPRMVPRAVYLPNCDRKGFTKRKCKPSGRRR 152  
OY 207 GYCWCVDEGGVKKVPGSD 223  
DB 153 GICWCVDKYGKMKLPQME 169

RESULT 6  
US-08-839-032A-6

Sequence 6, Application US/08839032A  
Patent No. 5891675  
GENERAL INFORMATION:  
APPLICANT: Lin, Lih-Ling  
APPLICANT: Chen, Jennifer H.  
APPLICANT: Schievella, Andrea  
APPLICANT: Graham, James  
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/839,032A  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A, 32,724  
REGISTRATION NUMBER: G15232DDIV  
REFERENCE/DOCKET NUMBER: G15232DDIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 186 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-839-032A-6

Query Match 9.5%; Score 122.5; DB 2; Length 186;  
Best Local Similarity 35.1%; Pred. No. 3.4e-05;  
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

OY 151 ELECTROOGRPTCEHLSCEERIKAHNSLRVEMFVPECLDGSYNPOQWPS---T 206  
DB 96 EMROESEOG--PCRRHM-EASLOELKASPRMVPRAVYLPNCDRKGFTKRKCKPSGRRR 152  
OY 207 GYCWCVDEGGVKKVPGSD 223  
DB 153 GICWCVDKYGKMKLPQME 169

RESULT 7  
US-08-839-031A-6  
Sequence 6, Application US/08839031A  
Patent No. 5948638

```

; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,031A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.,
; REGISTRATION NUMBER: 41,323
; REFERENCE/DOCKET NUMBER: G15232BDIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 186 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-839-031A-6

Query Match          9.5%; Score 122.5; DB 2; Length 186;
Best Local Similarity 35.1%; Pred. No. 3.4e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

Qy 151 ELECTROGRPTCEHLSCEEARIKAHNSLRLVEMFVECELEDGSYNPQCWPS---T 206
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 96 EMROSEEG--PCRHH-ESLQELKASPRMVPRAVYLLPCNDKRGFKRKCKRSRGR 152
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 207 GYCWCVDEGGVYKPGSD 223
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 153 GICWCVDKTKMKLPQME 169
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
PCT-US95-12724-6
; Sequence 6, Application PC/TUS9512724
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/12724
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.,
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15232B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 186 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-12724-6

Query Match          9.5%; Score 122.5; DB 5; Length 186;
Best Local Similarity 35.1%; Pred. No. 3.4e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

Qy 151 ELECTROGRPTCEHLSCEEARIKAHNSLRLVEMFVECELEDGSYNPQCWPS---T 206
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 96 EMROSEEG--PCRHH-ESLQELKASPRMVPRAVYLLPCNDKRGFKRKCKRSRGR 152
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 207 GYCWCVDEGGVYKPGSD 223
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 153 GICWCVDKTKMKLPQME 169
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
5212074-7
; Patent No. 5212074
; APPLICANT: KIEFER, MICHAEL C.; MASJAR, FRANK R.
; TITLE OF INVENTION: GENETIC MATERIAL ENCODING NEW
; INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN IGFBP-6
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/576,629
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 574,613
; FILING DATE: 28-AUG-1990
; SEQ ID NO: 7
; LENGTH: 237
; 5212074-7

Query Match          9.5%; Score 122.5; DB 6; Length 237;
Best Local Similarity 23.0%; Pred. No. 4.6e-05;
Matches 52; Conservative 30; Mismatches 73; Indels 71; Gaps 13;

Qy 37 COOLAASANSGL-----IGTVYPOCKETEFPEDEKCKWSTGYCVCVDEDEKEL 85
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 23 CEELVREPGCCATCALGICMPCGYTRRCG-----SLKCYTPRGV-----EKL 69
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 86 GTRKIRSPDCSRRAALTLCCOMQATITVVPWCGPSPCKADG-----SFDEVQCCASNG 140
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70 HTLMGGQVCMELAEIADIESIQ-----PSDKDGDHPNNSFS--PCSAHDR 115
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 141 ECQVDR-----KGKELGTRQGRP-----TCERHLSECEEARIKAHNSLR 183
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116 R-CLQKHAKITRDRSTSGGKMKVNGAPREDARVPYQSCQSELHRALE-RLAASGRTH 172
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 184 VEMF---VPECELDGSTRNPQCWPS---TGCKCYD-EGGVKYPG 221
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 173 EDLYITIPNCDRNGNFPKQCHALDGGRCWCVDKRTGVKLPG 218
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 10
US-08-698-551-8
; Sequence 8, Application US/08698551
; Patent No. 5712381
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/698,551
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-698-551-8

Query Match          9.5%; Score 122.5; DB 1; Length 272;
Best Local Similarity 35.1%; Pred. No. 5.6e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

QY 151 ELEGTRQGRPTCERHLSCEDEARIKANSLSRVEMFVPECLDGSYNPVQCWPS---T 206
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 182 EMROESBQG--PCRRHM-EASLQELKASPRMVPRAVYLPCNDKRGFKRKQCKPSRGRKR 238
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 207 GYCWCVDEGVKVPYPSD 223
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 239 GTCWCVDKYGKMLPGME 255

RESULT 11
US-08-602-228-8
; Sequence 8, Application US/08602228
; Patent No. 5843675
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
```

```
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,228
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-602-228-8

Query Match          9.5%; Score 122.5; DB 2; Length 272;
Best Local Similarity 35.1%; Pred. No. 5.6e-05;
Matches 27; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

QY 151 ELEGTRQGRPTCERHLSCEDEARIKANSLSRVEMFVPECLDGSYNPVQCWPS---T 206
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 182 EMROESBQG--PCRRHM-EASLQELKASPRMVPRAVYLPCNDKRGFKRKQCKPSRGRKR 238
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 207 GYCWCVDEGVKVPYPSD 223
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 239 GTCWCVDKYGKMLPGME 255

RESULT 12
US-08-649-341A-8
; Sequence 8, Application US/08649341A
; Patent No. 5847099
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,341A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-649-341A-8
```









PI Strukelj B, Turk V,  
XX



XX 02-APR-2001 (first entry)  
 DT  
 XX  
 DE Mouse TANGO 209 variant 1 polypeptide.  
 XX  
 KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;  
 KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;  
 KW antiasthmatic; neuroprotective; cytosolic; cardiant; hepatotropic;  
 KW antiinflammatory; antidiabetic; antinfertility; antipyretic; vasotropic;  
 KW antineumatic; nephrotropic; hemostatic; antileptic; osteopathic;  
 KW ophthalmological; antisticking; antileuc; vulnery; variant.  
 KW  
 XX  
 OS Mus sp.

XX Key Location/Qualifiers  
 FH MISC-difference 65 /label= E65D  
 FT /note= "Wild-type Glu is replaced by Asp"  
 XX  
 PN W0200069885-A2.

XX 23-NOV-2000.

XX 15-MAY-2000; 2000MO-US13361.

XX 14-MAY-1999; 99US-0312359.

XX (MILL-) MILLENNIUM PHARM INC.

XX Pan Y, Leiby KR;

XX WPI: 2001-024999/03.

XX N-PSDB: AAC84410.

XX Novel nucleic acids encoding secreted or transmembrane proteins, useful  
 PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of  
 PT the lung, liver, kidney or pancreas -  
 PS  
 XX

Claim 8; Page -: 20pp; English.

XX The invention provides human and mouse nucleic acids designated TANGO  
 CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane  
 CC proteins. The polypeptides, nucleic acids and their modulators may be  
 CC useful for treating or modulating cholesterol uptake, blood coagulation,  
 CC to modulate cell proliferation, morphogenesis and fate specification,  
 CC tissue repair and renewal, to treat cancer and promote wound healing,  
 CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan  
 CC syndrome, protein S deficiency, modulate allergic or inflammatory  
 CC response, acid secretion, tropic effects on gastrointestinal mucosa, and  
 CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,  
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,  
 CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain  
 CC herniations, meningitis, ischemic brain or heart disease, infarction,  
 CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart  
 CC disease, pulmonary heart disease, rheumatic fever, congenital heart  
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,  
 CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,  
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's  
 CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,  
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, Mcune  
 CC Albright syndrome, infertility, uterine disorders, viral disease. The  
 CC present sequence represents a mouse TANGO 209 variant polypeptide.  
 CC Note: the present variant sequence has been constructed using the  
 CC information provided in the specification.  
 CC  
 XX  
 SO Sequence 447 AA:

Query Match 20.2%; Score 259.5; DB 22; Length 447;

Best Local Similarity 28.9%; Pred. No. 1.3e-14;  
 Matches 36; Conservative 30; Mismatches 86; Indels 27; Gaps 5;

QY 41 QASANSGLIGTVPOCKTEFEFEKOCMGSTGYCWCVDDEKILGTIKI-RGSPDCSRK 99

DB 98 qeqarkefgvifpecdnddgtysgqchsygycwctvngprisgtaavahkprp---- 154  
 QY 100 AALTCCMMQALIVVPCWCG-----PPSCRADSFDEVCCASNGECYCDKKGAE 151  
 DB 155 -----gslnexkypqregagkaddaapaletpqgdeediasrptlwtegyksrq 205  
 QY 152 LEGRROGRPTCERHLSCEEARIRKANSLSRVEMFVPECTLEDGSYPVOCWPGTGYCWC 211  
 DB 206 ntknksascdqehgsaleekqpkndh-----vvlpeahgsllykpvqchpsrtygycw 260  
 QY 212 V-DEGGYKVPGSVPRKRTFC 231  
 DB 261 vlvdtgrlpipsttyeqpkc 281

RESULT 4  
 AAB48143  
 ID AAB48143 standard; Protein; 447 AA.

AC AAB48143;

DT 02-APR-2001 (first entry)

DE Mouse TANGO 209 variant 2 polypeptide.

XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;  
 KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;  
 KW antiasthmatic; neuroprotective; cytosolic; cardiant; hepatotropic;  
 KW antiinflammatory; antidiabetic; antinfertility; antipyretic; vasotropic;  
 KW antineumatic; nephrotropic; hemostatic; antileptic; osteopathic;  
 KW ophthalmological; antisticking; antileuc; vulnery; variant.  
 KW  
 XX  
 OS Mus sp.

XX Key Location/Qualifiers  
 FH MISC-difference 77 /label= E77D  
 FT /note= "Wild-type Glu is replaced by Asp"  
 XX  
 PN W0200069885-A2.

XX 23-NOV-2000.

XX 15-MAY-2000; 2000MO-US13361.

XX 14-MAY-1999; 99US-0312359.

XX (MILL-) MILLENNIUM PHARM INC.

XX Pan Y, Leiby KR;

XX WPI: 2001-024999/03.

XX N-PSDB: AAC84411.

XX Novel nucleic acids encoding secreted or transmembrane proteins, useful  
 PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of  
 PT the lung, liver, kidney or pancreas -  
 PS  
 XX

Claim 8; Page -: 20pp; English.

XX The invention provides human and mouse nucleic acids designated TANGO  
 CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane  
 CC proteins. The polypeptides, nucleic acids and their modulators may be  
 CC useful for treating or modulating cholesterol uptake, blood coagulation,  
 CC to modulate cell proliferation, morphogenesis and fate specification,  
 CC tissue repair and renewal, to treat cancer and promote wound healing,  
 CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan  
 CC syndrome, protein S deficiency, modulate allergic or inflammatory  
 CC response, acid secretion, tropic effects on gastrointestinal mucosa, and  
 CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,  
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,  
 CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain



ophthalmological; antislaking; antilicer; vulnerary.

Homo sapiens.

WO200069885-A2.

23-NOV-2000.

15-MAY-2000; 2000MO-US13361.

14-MAY-1999; 99US-0312359.

(MILL-) MILLENNIUM PHARM INC.

Pan Y, Leiby KR;

WPI: 2001-024999/03.

N-PSDB; AAC84381, AAC84382.

Novel nucleic acids encoding secreted or transmembrane proteins, useful for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of the lung, liver, kidney or pancreas -

Claim 8; Fig 14A-E; 209p; English.

The invention provides human and mouse nucleic acids designated TANGO 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane proteins. The polypeptides, nucleic acids and their modulators may be useful for treating or modulating cholesterol uptake, blood coagulation, to modulate cell proliferation, morphogenesis and fate specification, tissue repair and renewal, to treat cancer and promote wound healing, modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan syndrome, protein S deficiency, modulate allergic or inflammatory response, acid secretion, tropic effects on gastrointestinal mucosa, and promote ulcer healing, treat bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis, leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain herniations, meningitis, ischemic brain or heart disease, infarction, intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart disease, pulmonary heart disease, rheumatic fever, congenital heart disease, myocardial disease, atherosclerosis, hypertension, jaundice, hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome, sickle cell disease, renal failure, ischemic bowel disease, Crohn's disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome, CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune Albright syndrome, infertility, uterine disorders, viral disease. The present sequence represents the human TANGO 209 polypeptide.

Sequence 446 AA:

Query Match 19.8%; Score 254.5; DB 22; Length 446;

Best Local Similarity 28.2%; Pred. No. 3, 6e-14;

Matches 59; Conservative 29; Mismatches 78; Indels 43; Gaps 7;

41 QASANSGLIGTYVPOCKEETGEFEKOCWSTGVCWCDDEGEILGRT-RGSPDCRRK 99

98 gqgarketqyflpecdndgtysqychsytygcwcvupngpplsgtavakprc----- 153

100 AALTLGQMMAIIVNPGMGP--PSCKADGSPDEV-----OCCANGCNCYC 144

154 -----pysvneklpqreigtgtdaapaletqpgqdeasrptlw 137

145 VVKKKKELEGRROGRPTCEP-HLSECEARIKAHNSIRKEMVPECLLEGSGINPYQCV 203

198 teqvkrsqntknksvsscdqesaleekqpkndh-----vvlpecatgylkpvqch 252

204 PSTGYCMCV-DEGGKVVGSDVFRKRPCT 231

253 pstgycmcvlvtgtrpigtstsyepqkc 281

RESULT 7

AAB48139  
ID AAB48139 standard; Protein; 446 AA.

AAB48139;

02-APR-2001 (first entry)

Human TANGO 209 variant 1 polypeptide.

TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;

transmembrane protein; antineutropenic; cerebroprotective; atherosclerosis;

antistaphylococcal; neuroprotective; cytoprotective; hepatotropic;

antimicrobial; antidiabetic; antifertility; antipruritic; vasotrophic;

antirheumatic; nephroprotective; hemostatic; antileukemic; osteoprotective;

ophthalmological; antislaking; antilicer; vulnerary; variant.

Homo sapiens.

Key Location/Qualifiers

Misc-difference 65 /label= E65D

/note="wild-type Glu is replaced by Asp"

WO200069885-A2.

23-NOV-2000.

15-MAY-2000; 2000MO-US13361.

14-MAY-1999; 99US-0312359.

(MILL-) MILLENNIUM PHARM INC.

Pan Y, Leiby KR;

WPI: 2001-024999/03.

N-PSDB; AAC84407.

Novel nucleic acids encoding secreted or transmembrane proteins, useful for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of the lung, liver, kidney or pancreas -

Claim 8; Page 7; 209p; English.

The invention provides human and mouse nucleic acids designated TANGO 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane proteins. The polypeptides, nucleic acids and their modulators may be useful for treating or modulating cholesterol uptake, blood coagulation, to modulate cell proliferation, morphogenesis and fate specification, tissue repair and renewal, to treat cancer and promote wound healing, modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan syndrome, protein S deficiency, modulate allergic or inflammatory response, acid secretion, tropic effects on gastrointestinal mucosa, and promote ulcer healing, treat bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis, leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain herniations, meningitis, ischemic brain or heart disease, infarction, intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart disease, myocardial heart disease, atherosclerosis, hypertension, jaundice, hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome, sickle cell disease, renal failure, ischemic bowel disease, Crohn's disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome, neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune Albright syndrome, infertility, uterine disorders, viral disease. The present sequence represents a human TANGO 209 variant polypeptide. Note: the present variant sequence has been constructed using the information provided in the specification.

Sequence 446 AA:

Query Match 19.8%; Score 254.5; DB 22; Length 446;

The insertion provides human and mouse nucleic acids designated RANGO CC 204, RANGO 206, RANGO 209 and A236 encoding secreted or transmembrane CC proteins. The polypeptides, nucleic acids and their modulators may be useful for treating or modulating cholesterol uptake, blood coagulation, CC to modulate cell proliferation, morphogenesis and fate specification, CC tissue repair and renewal, to treat cancer and promote wound healing, CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan syndrome, protein S deficiency, modulate allergic or inflammatory

PR 14-MAY-1999; 99US-0312359

XX



PA (MILL-) MILLENNIUM PHARM INC.

XX Pan Y, Leiby KR;

DR WPI: 2001-024999/03.

DR N-PSDB: AAC84409.

XX Novel nucleic acids encoding secreted or transmembrane proteins, useful  
PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of  
PT the lung, liver, kidney or pancreas -

PS Claim 8; Page -: 209pp; English.

XX The invention provides human and mouse nucleic acids designated TANCO  
CC 204, TANCO 206, TANCO 209 and A236 encoding secreted or transmembrane  
CC proteins. The polypeptides, nucleic acids and their modulators may be  
CC useful for treating or modulating cholesterol uptake, blood coagulation,  
CC to modulate cell proliferation, morphogenesis and fate specification,  
CC tissue repair and renewal, to treat cancer and promote wound healing,  
CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan  
CC syndrome, protein S deficiency, modulate allergic or inflammatory  
CC response, acid secretion, tropic effects on gastrointestinal mucosa, and  
CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,  
CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,  
CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain  
CC infections, meningitis, ischemic brain or heart disease, infarction,  
CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart  
CC disease, pulmonary heart disease, rheumatic fever, congenital heart  
CC disease, myocardial infarction, atherosclerosis, hypertension, jaundice,  
CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,  
CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's  
CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,  
CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune  
CC Albright syndrome, infertility, uterine disorders, viral disease. The  
CC present sequence represents a human TANCO 209 variant polypeptide.  
CC Note: the present variant sequence has been constructed using the  
CC information provided in the specification.

XX Sequence 446 AA:

Query Match 19.8%; Score 254.5; DB 22; Length 446;

Best Local Similarity 28.2%; Pred. No. 3,6e-14;

Matches 59; Conservative 29; Mismatches 78; Indels 43; Gaps 7;

QY 41 OASANSGLIGTYVPOCKETGEFEKOCMSGYCWCVDDEGKEITGTF-RGSPCCSRK 99

DB 98 qeqarketqyvfipcedddqlysqvqchstygywcvtprngpysgltevahtkprc----- 153

QY 100 AALTLCOMQAIIVNPGMCGP--PSCKADGSPDEV-----QCCASNGECYC 144

DB 154 -----pssvnek1pqrqegtktdaaapaletqpgqdeedlastryplw 197

QY 145 VDKKKELEGTROOGRPTCR-HLSECEARIRAHNSLRVEMFPECELEDDSSYPVPCW 203

DB 198 teqvysygnktknksvsdcqdeqnsaleekqpkndn-----vvlpecahgilykprvch 252

QY 204 PSTGTGKCV-DEGKVKVGSQVRRKPTC 231

DB 253 pstgywcwlvdtgrlpjgtrstlyeqpkc 281

RESULT 10

AAB19394

XX AAB19394 standard; Protein; 435 AA.

XX AAB19394;

XX 06-MAR-2001 (first entry)

XX Amino acid sequence of a human secreted protein.

XX Secreted protein; platelet disorder; stem cell disorder; osteoporosis;

KW osteoarthritis; burn; incision; ulcer; periodontal disease; neuropathy;

KW nervous system disease; bone growth; cosmetic plastic surgery;

KW gut protection; gut regeneration; fibrosis; cancer;

KW bone marrow transplantation.

XX Homo sapiens.

XX WO200061755-A2.

XX 19-OCT-2000.

XX 10-APR-2000; 2000MO-US09555.

XX 09-APR-1999; 99US-0128574.

XX 20-AUG-1999; 99US-0150054.

XX (CHIR) CHIRON CORP.

XX Garcia PD;

XX WPI: 2000-665133/64.

XX N-PSDB: AAC61880.

XX Claim 14; Page 55; 74pp; English.

XX AAB19393-B19407 represent secreted human proteins. The secreted proteins  
CC are useful in assays to determine their biological activities. The  
CC proteins can also be used as biomarkers to identify tissues or cell  
CC types which express the proteins. The polynucleotide molecules can be  
CC used as biomarkers for tissues or chromosomes and to elicit immune  
CC responses. The proteins and antibodies are useful in diagnosis and  
CC treatment of diseases associated with altered expression of these  
CC proteins. The proteins are also useful for prevention or treatment of  
CC platelet disorders, stem cell disorders, osteoporosis or osteoarthritis,  
CC burns, incisions, ulcers, periodontal diseases, central and peripheral  
CC nervous system diseases and neuropathies, for healing fractured bones  
CC and to induce cartilage and/or bone growth in cosmetic plastic surgery.  
CC The proteins are also useful for gut protection or regeneration, for the  
CC treatment of lung or liver fibrosis, for stimulating blood cell  
CC generation in patients receiving cancer chemotherapy and for treatment  
CC of bone marrow transplantation patients.

XX Sequence 435 AA:

Query Match 19.6%; Score 252.5; DB 21; Length 435;

Best Local Similarity 31.7%; Pred. No. 5,3e-14;

Matches 70; Conservative 26; Mismatches 88; Indels 37; Gaps 8;

QY 31 EASTLRCO-----QOASANSGLIGTYVPOCKETGEFEKOCMSGYCWCVDDEGKEITG 86

DB 89 dagqskcrleraqlaqkqkprgevfipcedqstlyqchtygywcvtprngpysgltevahtkprc 148

QY 87 TKIRG-SPDCSRRAAALTLCOMQAIIVNPGMCGPSCAKAGS-----FDDVQ 134

DB 149 ssvgnktkprvcsvtdkplsg-----gsgtkddqskpvtmetqpyfdgde 195

QY 135 CCASNGECYCVDKKGKKELEGTROOGRP--TCERHLSCECEARIRAHNSLRVEMFPECE 191

DB 196 lkaplwlkhlvlkdsknlnlnrlnseksydcq-----erqsaleeqqnp-regvlpecc 251

QY 192 LEDGSYNVOCWPSGTGKCV-DEGKVKVGSQVRRKPTC 231

DB 252 apgilykprvqchgstgywcwlvdtgrlpjgtrstlympsc 292

RESULT 11

AAB56363

XX AAB56363 standard; Protein; 238 AA.

[illegible]

Db 180 chgksdcfvcvdkdrevqvgltgxparhhpc 209

RESULT 12

AAM25372

ID AAM25372 standard; Protein: 176 AA.

AC AAM25372;

XX 16-OCT-2001 (first entry)

DT

XX Human protein sequence SEQ ID NO:887.

DE

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW antinflammatory; antirheumatic; antiarthritis; immunosuppressive;  
KW antibacterial; endocrine; cardiant; central nervous system; vitruclide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiagregant; haemostatic; vulnery; antilucer; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder.

XX Homo sapiens.

XX MO200153455-A2.

PN 26-JUL-2001.

XX 22-DEC-2000; 2000MO-US35017.

PF 23-DEC-1999; 99US-0471275.

XX 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

PI WPI: 2001-457603/49.

DR N-PDB; AAM99313.

XX Isolated human polynucleotides encoding polypeptides, useful for the

PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

PS Claim 20; Page 196; 1217PP; English.

XX AAM99166 to AAM99904 encode the human proteins given in AAM25225 to

CC AAM25963. The proteins can have activities based on the tissues and

CC cells they are expressed in, such as: antinflammatory; antirheumatic;  
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
CC central nervous system; vitruclide; anti-HIV; fungicide; antimutagen;  
CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnery;  
CC antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;  
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
CC encoding them can be used in gene therapy, antisense therapy and vaccine  
CC production. The proteins and polynucleotides are useful for screening for  
CC agonists or antagonists of a protein and for the treatment and diagnosis  
CC of disorders associated with the activity of a protein e.g. inflammation,  
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,



AA The present invention describes secreted and transmembrane polypeptides  
 CC and their polynucleotides. The nucleotide sequences are useful as  
 CC  
 CC



7  
:  
:  
: